

**LINKING MICROBIAL METABOLISM AND ORGANIC MATTER CYCLING
THROUGH METABOLITE DISTRIBUTIONS IN THE OCEAN**

By

Winifred M. Johnson

B.A., Haverford College, 2009

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

and the

WOODS HOLE OCEANOGRAPHIC INSTITUTION

February 2017

© 2017 Winifred M. Johnson

All rights reserved.

The author hereby grants to MIT and WHOI permission to reproduce and
to distribute publicly paper and electronic copies of this thesis document
in whole or in part in any medium now known or hereafter created.

Signature of Author _____
Joint Program in Oceanography/Applied Ocean Science and Engineering
Massachusetts Institute of Technology
and Woods Hole Oceanographic Institution
January 18, 2017

Certified by _____
Elizabeth B. Kujawinski
Thesis Supervisor

Accepted by _____
Shuhei Ono
Chair, Joint Committee for Chemical Oceanography
Woods Hole Oceanographic Institution

Linking microbial metabolism and organic matter cycling through metabolite distributions in the ocean

by

Winifred M. Johnson

Submitted to the MIT/WHOI Joint Program in Chemical Oceanography on January 18, 2017 in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Abstract

Key players in the marine carbon cycle are the ocean-dwelling microbes that fix, remineralize, and transform organic matter. Many of the small organic molecules in the marine carbon pool have not been well characterized and their roles in microbial physiology, ecological interactions, and carbon cycling remain largely unknown. In this dissertation metabolomics techniques were developed and used to profile and quantify a suite of metabolites in the field and in laboratory experiments. Experiments were run to study the way a specific metabolite can influence microbial metabolite output and potentially processing of organic matter. Specifically, the metabolic response of the heterotrophic marine bacterium, *Ruegeria pomeroyi*, to the algal metabolite dimethylsulfoniopropionate (DMSP) was analyzed using targeted and untargeted metabolomics. The manner in which DMSP causes *R. pomeroyi* to modify its biochemical pathways suggests anticipation by *R. pomeroyi* of phytoplankton-derived nutrients and higher microbial density. Targeted metabolomics was used to characterize the latitudinal and vertical distributions of particulate and dissolved metabolites in samples gathered along a transect in the Western Atlantic Ocean. The assembled dataset indicates that, while many metabolite distributions co-vary with biomass abundance, other metabolites show distributions that suggest abiotic, species specific, or metabolic controls on their variability. On sinking particles in the South Atlantic portion of the transect, metabolites possibly derived from degradation of organic matter increase and phytoplankton-derived metabolites decrease. This work highlights the role DMSP plays in the metabolic response of a bacterium to the environment and reveals unexpected ways metabolite abundances vary between ocean regions and are transformed on sinking particles. Further metabolomics studies of the global distributions and interactions of marine biomolecules promise to provide new insights into microbial processes and metabolite cycling.

Thesis supervisor: Dr. Elizabeth B. Kujawinski

Title: Associate Scientist with Tenure, WHOI

Acknowledgments

First, and foremost, I would like to thank my advisor, Liz Kujawinski for giving me the incredible opportunity to work with her on such exciting research. Her vision and determination to pursue this compelling but challenging work has been inspiring and allowed me to find my ideal scientific niche linking chemistry, ecology, and the environment. She has been a thoughtful and kind mentor, always making herself available to discuss research and provide guidance. I would also like to thank my committee members Phil Gschwend, Ben Van Mooy, and Colleen Hansel, as well as my defense chair, Mak Saito. Phil has pushed me to carefully consider the molecular structure of metabolites and how this relates to their analysis and their role in the ocean. Ben's ability to see the larger oceanographic context has helped me to put my data into perspective. Colleen's experience outside of organic matter and oceanography has brought interesting connections and possibilities to bear on my work.

All of the members of the Kujawinski Lab over my time here have taught me and informed my research. In particular, Melissa Kido Soule has patiently trained me to use the FTI-ICR and triple quadrupole mass spectrometers, and answered all of my questions about the inner workings of these instruments. Her ability to calmly and methodically break down a problem to find a solution has helped me through moments of panic. Krista Longnecker has answered so many of my questions over the years. I learned how to code by working off of scripts she shared with me and I have sought her advice on statistics on many occasions. I have also learned how to prepare for and conduct fieldwork at sea from Krista and appreciate her enthusiasm for going to sea. Both Melissa and Krista have provided meticulous and thoughtful feedback on manuscripts as well. I have also enjoyed working with and learning from Cara Fiore, Yina Liu, Maya Bhatia, Dave Griffith, Jacquie Carozza, Catherine Carmichael, Harriet Alexander, Bill Arnold, Gretchen Swarr, Crystal Brier, and Craig McLean.

This thesis would not have been possible without two research cruises, one on the R/V *Knorr* and the other on the R/V *Atlantic Explorer*. Thank you to the captains and crew of these ships. Mike Lomas, chief scientist on the R/V *Atlantic Explorer* cruise, kindly allowed me to join the cruise, collect samples from depths no one else on the ship wished to sample, and has since provided datasets to allow me to put the metabolite data into oceanographic context. I also thank all of the scientists who participated in these cruises for their assistance in sampling and making extended time at sea together enjoyable. In particular, Justin Ossolinski helped me with shipping and logistics, and Gwenn Hennon, Erin Eggleston, and Colleen Durkin provided me with data that they collected.

The community of scientists in Fye and at WHOI have been a source of assistance and friendship. Additionally, all of the administrative assistants in Fye and everyone in the Academic Programs Office have made this incredible experience possible. My fellow graduate students have been a great source of support, willing to talk through research problems, make coffee runs, and listen to practice talks. My fellow chemistry graduate students, including Jamie Collins, Nick Hawco, Emily Estes, Sarah Rosengard, Sophie Chu, Cara Manning, Kyrstin Fornace, Rene Boiteau, Evan Howard, Simone Moos, Tristan Kading, David Wang, and Jordon Hemingway were great study partners and friends throughout. Everyone who I shared housing with at Winding Lane and my friends throughout my time in the Joint Program have made these years

wonderful, especially Isabela Le Bras, Anna Wargula, Julie van der Hoop, Katie Pitz, and Oscar Sosa. I would also like to thank my parents and sister for their continued support and their help exploring the Cape each summer.

None of this work would have been possible without a variety of funding sources. I was supported for three years by a National Defense Science & Engineering Graduate Fellowship. The research was carried out with grants from the National Science Foundation (OCE-0928424 to EBK, OCE-1154320 to EBK and KL), the Gordon and Betty Moore Foundation (GBMF3304), Simons Foundation International, and the WHOI Ocean Ventures Fund.

Table of Contents

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| Chapter 1. Introduction | 9 |
| Chapter 2. Extraction efficiency and quantification of dissolved metabolites in targeted marine metabolomics | 23 |
| Abstract | 24 |
| 2.1 Introduction | 25 |
| 2.2 Materials and methods | 29 |
| 2.3 Assessment | 35 |
| 2.4 Discussion | 41 |
| 2.5 Conclusions | 42 |
| Acknowledgments | 43 |
| References | 44 |
| Supplemental Information | 55 |
| Chapter 3. Evidence for quorum sensing and differential metabolite production by a marine bacterium in response to DMSP | 79 |
| Abstract | 80 |
| 3.1 Introduction | 81 |
| 3.2 Materials and Methods | 85 |
| 3.3 Results | 87 |
| 3.4 Discussion | 91 |
| 3.5 Conclusions | 98 |
| Acknowledgments | 98 |
| References | 100 |
| Supplemental Information | 109 |
| Chapter 4. Insights into the controls on metabolite distributions along a latitudinal transect of the western Atlantic Ocean | 129 |
| Abstract | 130 |
| 4.1. Introduction | 131 |
| 4.2 Materials and methods | 134 |
| 4.3 Results | 140 |
| 4.4 Discussion | 148 |
| 4.5 Conclusions | 157 |
| Acknowledgments | 158 |
| References | 159 |
| Supplemental Information | 174 |
| Chapter 5. Metabolite composition of sinking particles reflects a changing microbial community and differential metabolite degradation | 183 |
| Abstract | 184 |
| 5.1 Introduction | 185 |
| 5.2 Materials and methods | 188 |
| 5.3 Results | 193 |
| 5.4 Discussion | 198 |
| 5.5 Conclusions | 206 |
| Acknowledgments | 207 |

| | |
|---------------------------------------|------------|
| References..... | 208 |
| Supplemental Information | 218 |
| Chapter 6. Conclusions..... | 227 |

Chapter 1. Introduction

1.1 Overview

Marine microbes are responsible for both fixation of carbon dioxide as well as its remineralization and transformation through the marine microbial food web (Azam et al. 1983). Approximately 50% of the carbon fixed by phytoplankton is cycled through the pool of dissolved organic matter (DOM) (Ducklow 1999), which is a carbon reservoir similar in size to that of the atmosphere. Some organic matter produced in the surface ocean sinks as aggregates, as part of dying phytoplankton cells, or as zooplankton fecal pellets and is subsequently remineralized by bacteria deep in the ocean or is buried in sediments (Volk and Hoffert 1985). Marine organisms play essential roles in all of these processes. The movement of organic molecules through microbial cells and then into other parts of the ecosystem defines, in large part, the ocean's essential and dynamic role in the global carbon cycle. By measuring these molecules within cells and suspended particles in the environment, dissolved in seawater, and as part of particles sinking through the water column, we can explore microbial communities' adaptations to varied ocean environments and trace the microbial processes that largely shape the ocean carbon cycle and ultimately climatic conditions on our planet. In this thesis, an approach called metabolomics is used to yield insights not only into the ecophysiology of marine organisms but also to directly link their metabolism to the cycling of organic matter in the ocean.

1.2 Metabolites reflect cellular phenotype

Metabolomics characterizes a suite of small organic molecules, or metabolites, to ascertain a cell, tissue, or organism's physiological state. Metabolites are the downstream products of gene transcription and enzyme activity and reflect a cell or organism's phenotype (Oliver et al. 1998; Tweeddale et al. 1998; Fiehn 2002). Specifically, metabolites include nucleic

acid precursors, amino acids, lipids, sugars, vitamins, biosynthetic intermediates, signaling molecules, and other secondary metabolites (for example, antibiotics). The simultaneous characterization of these molecules reveals the combined impact of a cell's genetic blueprints and environment on the metabolic functions of the cell.

Metabolite abundances can be controlled at many points in the cellular regulatory apparatus. While transcription and translation contribute to the regulation of metabolite abundances, metabolites themselves can also regulate gene transcription and enzyme activity. Enzyme activity can respond rapidly to changes in metabolite concentration through allosteric regulation. This occurs when a metabolite binds to an enzyme and causes a conformational change in the protein that either enhances or inhibits the enzyme's function (reviewed by Chubukov et al. 2014). In addition, metabolites can control gene transcription by binding to riboswitches. Riboswitches are non-coding regions of mRNA that activate or block transcription through structural changes induced by metabolite binding (reviewed by Mandal and Breaker 2004). Thus, metabolite abundances play a role in complex metabolic feedback loops that can control pathway activity and metabolite production.

Additionally, characterization of the metabolome can enhance understanding of the metabolic capabilities of an organism. Genome annotations can be inaccurate when using a gene sequence identified in one species to annotate a gene in a distantly related species. In particular, errors occur when assigning a function to enzymes that have similar structural motifs but not necessarily the same function (Schnoes et al. 2009). There are also genes that cannot be annotated because their function is simply unknown. In bacteria, approximately 30% of the proteins encoded in the genome have an unknown function (Meier et al. 2013), while in eukaryotes this value climbs to 40% (Horan et al. 2008). Thus, profiling of metabolites can

uncover novel pathways in organisms because prior knowledge of gene or enzyme function is not required to analyze the metabolome. Specific approaches such as stable isotope labeling to examine metabolite fluxes, using mutations to disrupt a pathway to examine the impact on the metabolome, or coupling metabolomics with transcriptomics have been used to elucidate new biochemical pathways (reviewed by Prosser et al. 2014).

1.3 Organic molecules in the ocean

Small organic molecules, as well as the biopolymers that they form, have been characterized in both particulate and dissolved samples in the ocean for decades. Analytical approaches have primarily focused on individual classes of molecules or even on individual molecules as specific extraction and analytical techniques were required for detection and quantification of these molecules. Major classes of biomolecules that have been studied include amino acids, carbohydrates, and lipids. Additionally, less quantitatively dominant organic molecules have also been studied on an individual basis or in small groups.

Amino acids and proteins comprise 40-50% of the particulate organic carbon (by weight) measured in the upper ocean (Hedges et al. 2001). Individual dissolved free amino acids were found to show diel variability with maximum concentrations in the evening and minimum concentrations early in the morning. These observations imply heterotrophic uptake of these molecules (Mopper and Lindroth 1982). While free amino acids are essential metabolites for exchange between marine microbes, total hydrolysable amino acids (i.e., both protein-bound and free amino acids) have been more extensively characterized due to the greater abundance of protein incorporated amino acids, and their importance as a pool of organic nitrogen (Lee and Cronin 1982; Dittmar et al. 2001). In addition, relative amino acid abundances have been found to change with increasing organic matter degradation, thus leading to the development of amino

acid-based degradation indices (Dauwe et al. 1999).

Monomeric carbohydrates as well as polysaccharides are synthesized by marine organisms and are important sources of energy. Phytoplankton, such as coccolithophores can release large amounts of dissolved carbohydrates with these molecules increasing over the growth cycle ranging from 26-80% of the total organic carbon measured (Biddanda and Benner 1997). Carbohydrates are also components of sinking particles comprising around 30% (by mass) and becoming a slightly larger proportion with depth (Hedges et al. 2001).

Lipids are another major class of biomolecules that have been characterized in the ocean. In particular, selective degradation or preservation of certain types of lipids have been observed in sinking particles (Wakeham et al. 1997). Fatty acid lipids are highly abundant lipids in all types of marine microbes but different structural types are derived from phytoplankton versus heterotrophic bacteria and all are rapidly degraded in the surface ocean (Van Mooy and Fredricks 2010).

Past research has also focused on low abundance biological molecules that play important roles in microbial physiology and microbial communities. For instance B vitamins have previously been characterized in the ocean, with a particular focus on vitamin B₁₂ due to its potential to co-limit phytoplankton productivity (Ohwada and Taga 1972; Bertrand et al. 2007; Sañudo-Wilhelmy et al. 2012; Heal et al. 2016). Osmolytes, which regulate the osmotic pressure of cells, and are thus present at high concentrations in the cytosol are also of interest as organic substrates for other microbes. For example, dimethylsulfoniopropionate (DMSP), is an organosulfur metabolite produced by phytoplankton and its prevalence in the ocean can be linked to phytoplankton abundance and stochastic events such as blooms (Kiene and Linn 2000; Levine et al. 2012). Other osmolytes like betaine, a nitrogen-containing osmolyte, have been measured

in certain phytoplankton species (Keller et al. 1999). Thiols, such as glutathione, and other combinations of amino acids containing cysteine can protect cells from the toxic effects of cadmium and copper. These molecules have been quantified in both particulate and dissolved pools, with glutathione generally being the largest portion of the thiol pool (Ahner et al. 2002; Dupont et al. 2006).

While the compound-class based approach to characterizing marine organic matter is valuable, it does not capture holistic changes in multiple areas of metabolism simultaneously. These responses are difficult to predict due to the complex networks of biochemical pathways inside of cells. In addition, focus on specific classes of molecules constrains work to known molecules. By taking a metabolomics approach to studying these molecules, the aim is to characterize how the physiological state of marine microbes relates to the composition of the molecules that are released into the marine food web and how this varies across the ocean. By applying metabolomics in the ocean, we hope to uncover novel physiological adaptations, microbial interactions, and how the exchange and cycling of these small organic molecules links to global geochemical cycles.

1.4 Analytical approaches to metabolomics

To analyze metabolites in seawater, they must be separated from the salt and concentrated so that they can be detected on the instrument. The comprehensive extraction of the metabolome presents challenges due to the structural diversity of the molecules. Solid phase extraction, where polymer-molecule interactions determine the extraction of metabolites from their matrix, is an effective way to concentrate metabolites and remove salt, but each resin type selects for certain molecular structures. Currently, there is no resin that can comprehensively extract metabolites from seawater. However, the PPL polymer that has been found to extract

~50% of dissolved marine organic matter (Dittmar et al. 2008), and is most frequently used to extract small organic molecules from seawater.

Metabolomics requires cutting-edge technologies, usually involving nuclear magnetic resonance or mass spectrometry, to resolve individual molecules within a complex mixture. Liquid chromatography-mass spectrometry (LC-MS) separates metabolites by polarity on a column and then generates a mass-to-charge ratio of individual ions. In this thesis, both Fourier transform ion cyclotron resonance (FT-ICR) MS and triple-stage quadrupole (QqQ) MS are used for the analyses. FT-ICR-MS is used for untargeted metabolomics to resolve thousands of molecules. This instrument is coupled to a linear ion trap that allows ions to be captured and fragmented to aid in the identification of unknown molecules. In contrast, QqQ-MS is used for targeted metabolomics and selects for pre-designated parent ions and then detects fragments of those ions. The samples can then be compared to calibration curves for identification confirmation and quantification.

Combining these two types of analysis has the potential to advance discovery of important novel or unanticipated metabolites. In particular, a metabolite quantified using the targeted method can support the identification of metabolites derived from the same biochemical pathway that are measured in the untargeted method and show similar responses or patterns. Conversely, a metabolite can be putatively identified in the untargeted method and then added to the targeted method to confirm its identity and concentration.

1.5 Development of metabolomics

The term metabolomics evolved out of the development of other “omics” techniques. It was initially applied to single-cell model organisms before being expanded to biomedical applications. Additionally, it has been used in different ecosystems to elucidate questions of

physiological adaptations and species interactions. One of the first papers to discuss the need for metabolomics examined the fully sequenced yeast genome and the 4000 new protein-encoding genes whose function now needed to be understood (Oliver et al. 1998). Oliver et al. (1998) proposed targeted gene knockout studies to generate metabolic profile changes which would indicate the function of genes. In that vein, an early metabolomics study in *Escherichia coli* used gene knockout studies to identify areas of metabolism not known to be involved in adaptation to nutrient limitation (Tweeddale et al. 1998). These studies led the way to biomedical research that could compare healthy and non-healthy individuals to diagnose and identify mechanisms of a disease.

In medicine, the metabolite profiles of blood, urine, or saliva are often characterized. These studies have led to a better understanding of diseases through identification of disease metabolite biomarkers. These investigations have also expanded our knowledge of the role of our own microbiome in human health through identification of metabolites produced by gut microbiota rather than by human cells. Microbes in the gut can produce amino acids and their derivatives that enter the bloodstream and can provide biochemical services to the host, such as indole-3-propionic acid acting as an antioxidant (Wikoff et al. 2009). In studies of a drug for treating diabetes, decreased abundance of the metabolites, citrate and hippurate, was observed in individuals receiving treatment. Elevated concentrations of these metabolites had previously been identified in individuals with diabetes, suggesting that a beneficial response to the drug could be observed in the metabolome (Van Doorn et al. 2007). Metabolite biomarkers for many types of diseases are actively sought using metabolomics. For instance, there is hope that identification of biomarkers for certain cancers may aid in earlier detection, diagnosis, and treatment (reviewed by Wang et al. 2013; Armitage and Barbas 2014).

In the environment, metabolomics can be used to study the physiological adaptations of organisms to environmental conditions (Bundy et al. 2009). In a number of species, the response of the metabolome to environmental inputs has been examined. These include factors such as sudden temperature changes (Kaplan et al. 2004) or nutrient limitation (Warne et al. 2001; Brauer et al. 2006), as well as other abiotic changes in the environment and both beneficial or antagonistic interactions with other organisms. These shifts in metabolite composition can be indicative of the biochemical mechanisms used to respond to the environment.

In addition, environmental metabolomics is expanding our understanding of the relationship between organisms and their environment and their role in elemental cycling. In soil, metabolomics can be used to trace molecules being degraded and produced by soil microbes (Swenson et al. 2015). Metabolomics can also be used to assess ecosystem health and recovery by measuring microbial incorporation of organic pollutants (Jones et al. 2014). Finally, it can add to our understanding of organismal interactions, for example, the essential role bacteria play in the rhizosphere of plants (Narasimhan et al. 2003; Barret et al. 2013).

Throughout the fields in which metabolomics has been applied, it has been a valuable tool to discover new pathways, genes, and adaptations. It has also facilitated the tracking of chemical interactions between organisms and the subsequent responses, whether that be between a human and their gut microbiome or within a plant-microbe mutualism. Metabolomics holds promise for the discovery of new diagnostic metabolites for disease and treatment but also for ecosystem health and recovery. Finally, in the environment, metabolites form a direct link between biological processes and elemental cycling, which is a developing area of research in both terrestrial and aquatic systems.

1.6 Metabolomics in the ocean

In the marine environment, thus far, metabolomics has been predominantly applied to species that can be cultured under controlled laboratory conditions. Studies using untargeted metabolomics with cultures of marine microbes have demonstrated that metabolites vary with growth stage in a marine diatom (Barofsky et al. 2009) and that many metabolites produced cannot be predicted based on genomic information in a *Synechococcus* species (Baran et al. 2010). The fluctuation of different osmolytes in phytoplankton species as a function of salinity has also been observed (Gebser and Pohnert 2013). Co-culturing experiments have demonstrated that bacteria can induce changes in the metabolome of phytoplankton (Paul et al. 2012) and that viral infection can also lead to increased production of certain metabolites (Ankrah et al. 2014; Rosenwasser et al. 2014). Culture experiments have also identified non-enzymatic production of metabolites through reactions with intracellular reactive oxygen species in *Synechococcus elongatus* CCMP 1631 (Fiore et al. 2015) and putatively identified compounds not previously associated with *Thalassiosira pseudonana* (Longnecker et al. 2015). These studies highlight the extent to which metabolite composition can vary within a species in response to a single experimental perturbation. However, these controlled laboratory experiments only hint at the complex environmental information reflected in marine microbial *in situ* phenotypes.

The great variety of factors that can shape metabolite composition must continue to be explored in an array of species. In addition, these studies must begin to elucidate the mechanisms that control metabolite abundances to create a foundation upon which metabolite profiles in the field can be interpreted. In addition, tracing these molecules through the food web and the environment as they are exchanged and modified may reveal unknown species interactions. Notably, up to this point marine metabolomics has been confined to controlled laboratory experiments, which are essential. However, in tandem, the distributions and variability of

metabolites in the ocean need to be understood to generate hypotheses and provide context to laboratory studies.

1.7 Summary

This thesis both advances our technical ability to study metabolites in the ocean and characterizes some of the first targeted metabolite profiles in the ocean, providing a foundation for future metabolomics studies in the marine system. Chapter 2 characterizes the extraction efficiency and impact of matrix effects on targeted metabolite analysis in marine samples, enhancing our ability to make these measurements in the ocean. Chapter 3 examines how a heterotrophic marine bacterium responds to an algal metabolite, DMSP, in laboratory culture, illustrating the impact of metabolite exchange on downstream metabolism and metabolite production. Chapters 4 and 5 examine metabolite variability in a variety of oceanic regions and on the sinking particles that deliver organic matter to depth. Characterization of marine metabolites in this thesis allows us to track how marine microbes respond to, produce, and transform metabolites, to gain a more detailed, molecule-specific understanding of the interaction between microbial life and organic matter cycling in the ocean.

References

- Ahner, B., L. Wei, J. Oleson, and N. Ogura. 2002. Glutathione and other low molecular weight thiols in marine phytoplankton under metal stress. *Mar. Ecol. Prog. Ser.* **232**: 93–103.
- Ankrah, N. Y. D., A. L. May, J. L. Middleton, and others. 2014. Phage infection of an environmentally relevant marine bacterium alters host metabolism and lysate composition. *ISME J.* **8**: 1089–1100. doi:10.1038/ismej.2013.216
- Armitage, E. G., and C. Barbas. 2014. Metabolomics in cancer biomarker discovery: current trends and future perspectives. *J. Pharm. Biomed. Anal.* **87**: 1–11. doi:10.1016/j.jpba.2013.08.041
- Azam, F., T. Fenchel, J. G. Field, L. A. Meyer-Reil, and F. Thingstad. 1983. The ecological role of water-column microbes in the sea. *Mar. Ecol. Prog. Ser.* **10**: 257–263.
- Baran, R., B. P. Bowen, N. J. Bouskill, E. L. Brodie, S. M. Yannone, and T. R. Northen. 2010. Metabolite identification in *Synechococcus* sp. PCC 7002 using untargeted stable isotope

- assisted metabolite profiling. *Anal. Chem.* **82**: 9034–9042. doi:10.1021/ac1020112
- Barofsky, A., C. Vidoudez, and G. Pohnert. 2009. Metabolic profiling reveals growth stage variability in diatom exudates. *Limnol. Oceanogr. Methods* **7**: 382–390.
- Barret, M., H. Tan, F. Egan, J. P. Morrissey, J. Reen, and F. O’Gara. 2013. Exploiting new systems-based strategies to elucidate plant-bacterial interactions in the rhizosphere. *Mol. Microb. Ecol. Rhizosph.* **1**: 57–68. doi:10.1002/9781118297674.ch6
- Bertrand, E. M., M. a. Saito, J. M. Rose, C. R. Riesselman, M. C. Lohan, A. E. Noble, P. a. Lee, and G. R. DiTullio. 2007. Vitamin B12 and iron colimitation of phytoplankton growth in the Ross Sea. *Limnol. Oceanogr.* **52**: 1079–1093. doi:10.4319/lo.2007.52.3.1079
- Biddanda, B., and R. Benner. 1997. Carbon, nitrogen, and carbohydrate fluxes during the production of particulate and dissolved organic matter by marine phytoplankton. *Limnol. Oceanogr.* **42**: 506–518. doi:10.4319/lo.1997.42.3.0506
- Brauer, M. J., J. Yuan, B. D. Bennett, W. Lu, E. Kimball, D. Botstein, and J. D. Rabinowitz. 2006. Conservation of the metabolomic response to starvation across two divergent microbes. *Proc. Natl. Acad. Sci. U. S. A.* **103**: 19302–19307. doi:10.1073/pnas.0609508103
- Bundy, J. G., M. P. Davey, and M. R. Viant. 2009. Environmental metabolomics: A critical review and future perspectives. *Metabolomics* **5**: 3–21. doi:10.1007/s11306-008-0152-0
- Chubukov, V., L. Gerosa, K. Kochanowski, and U. Sauer. 2014. Coordination of microbial metabolism. *Nat. Rev. Microbiol.* **12**: 327–340. doi:10.1038/nrmicro3238
- Dauwe, B., J. J. Middelburg, P. M. J. Herman, and C. H. R. Heip. 1999. Linking diagenetic alteration of amino acids and bulk organic matter reactivity. *Limnol. Oceanogr.* **44**: 1809–1814. doi:10.4319/lo.1999.44.7.1809
- Dittmar, T., H. Fitznar, and G. Kattner. 2001. Origin and biogeochemical cycling of organic nitrogen in the eastern Arctic Ocean as evident from D- and L-amino acids. *Geochim. Cosmochim. Acta* **65**: 4103–4114.
- Dittmar, T., B. Koch, N. Hertkorn, and G. Kattner. 2008. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol. Oceanogr. Methods* **6**: 230–235. doi:10.4319/lom.2008.6.230
- Van Doorn, M., J. Vogels, A. Tas, E. J. Van Hoogdalem, J. Burggraaf, A. Cohen, and J. Van Der Greef. 2007. Evaluation of metabolite profiles as biomarkers for the pharmacological effects of thiazolidinediones in type 2 diabetes mellitus patients and healthy volunteers. *Br. J. Clin. Pharmacol.* **63**: 562–574. doi:10.1111/j.1365-2125.2006.02816.x
- Ducklow, H. W. 1999. Minireview: The bacterial content of the oceanic euphotic zone. *FEMS Microbiol.* **30**: 1–10. doi:10.1016/S0168-6496(99)00031-8
- Dupont, C. L., J. W. Moffett, R. R. Bidigare, and B. A. Ahner. 2006. Distributions of dissolved and particulate biogenic thiols in the subarctic Pacific Ocean. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **53**: 1961–1974. doi:10.1016/j.dsr.2006.09.003
- Fiehn, O. 2002. Metabolomics—the link between genotypes and phenotypes. *Plant Mol. Biol.* **48**: 155–71.
- Fiore, C. L., K. Longnecker, M. C. Kido Soule, and E. B. Kujawinski. 2015. Release of ecologically relevant metabolites by the cyanobacterium, *Synechococcus elongatus* CCMP 1631. *Environ. Microbiol.* **17**: 3949–3963. doi:10.1111/1462-2920.12899
- Gebser, B., and G. Pohnert. 2013. Synchronized regulation of different zwitterionic metabolites in the osmoadaptation of phytoplankton. *Mar. Drugs* **11**: 2168–82. doi:10.3390/md11062168
- Heal, K. R., W. Qin, F. Ribalet, and others. 2016. Two distinct pools of B₁₂ analogs reveal community interdependencies in the ocean. *Proc. Natl. Acad. Sci.* **114**: 201608462.

- doi:10.1073/pnas.1608462114
- Hedges, J. I., J. A. Baldock, Y. Gélinas, C. Lee, M. Peterson, and S. G. Wakeham. 2001. Evidence for non-selective preservation of organic matter in sinking marine particles. *Nature* **409**: 801–804. doi:10.1038/35057247
- Horan, K., C. Jang, J. Bailey-Serres, and others. 2008. Annotating genes of known and unknown function by large-scale coexpression analysis. *Plant Physiol.* **147**: 41–57. doi:10.1104/pp.108.117366
- Jones, O. A. H., S. Sdepanian, S. Lofts, C. Svendsen, D. J. Spurgeon, M. L. Maguire, and J. L. Griffin. 2014. Metabolomic analysis of soil communities can be used for pollution assessment. *Environ. Toxicol. Chem.* **33**: 61–64. doi:10.1002/etc.2418
- Kaplan, F., J. Kopka, D. W. Haskell, and others. 2004. Exploring the temperature-stress metabolome. *Plant Physiol.* **136**: 4159–4168. doi:10.1104/pp.104.052142.1
- Keller, M. D., R. P. Kiene, P. A. Matrai, and W. K. Bellows. 1999. Production of glycine betaine and dimethylsulfoniopropionate in marine phytoplankton. I. Batch cultures. *Mar. Biol.* **135**: 237–248. doi:10.1007/s002270050621
- Kiene, R. P., and L. J. Linn. 2000. Distribution and turnover of dissolved DMSP and its relationship with bacterial production and dimethylsulfide in the Gulf of Mexico. *Limnol. Oceanogr.* **45**: 849–861. doi:10.4319/lo.2000.45.4.0849
- Lee, C., and C. Cronin. 1982. The vertical flux of particulate organic nitrogen in the sea: decomposition of amino acids in the Peru upwelling area and the equatorial Atlantic. *J. Mar. Res.* **40**: 227–251.
- Levine, N. M., V. A. Varaljay, D. A. Toole, J. W. H. Dacey, S. C. Doney, and M. A. Moran. 2012. Environmental, biochemical and genetic drivers of DMSP degradation and DMS production in the Sargasso Sea. *Environ. Microbiol.* **14**: 1210–1223. doi:10.1111/j.1462-2920.2012.02700.x
- Longnecker, K., M. C. Kido Soule, and E. B. Kujawinski. 2015. Dissolved organic matter produced by *Thalassiosira pseudonana*. *Mar. Chem.* **168**: 114–123. doi:10.1016/j.marchem.2014.11.003
- Mandal, M., and R. R. Breaker. 2004. Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* **5**: 451–463. doi:10.1038/nrm1403
- Meier, M., R. V. Sit, and S. R. Quake. 2013. Proteome-wide protein interaction measurements of bacterial proteins of unknown function. *Proc. Natl. Acad. Sci. U. S. A.* **110**: 477–82. doi:10.1073/pnas.1210634110
- Van Mooy, B. A. S., and H. F. Fredricks. 2010. Bacterial and eukaryotic intact polar lipids in the eastern subtropical South Pacific: Water-column distribution, planktonic sources, and fatty acid composition. *Geochim. Cosmochim. Acta* **74**: 6499–6516. doi:10.1016/j.gca.2010.08.026
- Mopper, K., and P. Lindroth. 1982. Diel and depth variations in dissolved free amino acids and ammonium in the Baltic Sea determined by shipboard HPLC analysis. *Limnol. Oceanogr.* **27**: 336–347. doi:10.4319/lo.1982.27.2.0336
- Narasimhan, K., C. Basheer, V. B. Bajic, and S. Swarup. 2003. Enhancement of plant-microbe interactions using a rhizosphere metabolomics-driven polychlorinated biphenyls. *Plant Physiol.* **132**: 146–153. doi:10.1104/pp.102.016295.populations
- Ohwada, K., and N. Taga. 1972. Distribution and seasonal variation of vitamin B12, thiamine and biotin in the sea. *Mar. Chem.* **1**: 61–73. doi:10.1016/0304-4203(72)90007-2
- Oliver, S. G., M. K. Winson, D. B. Kell, and F. Baganz. 1998. Systematic functional analysis of

- the yeast genome. *Trends Biotechnol.* **16**: 373–378.
- Paul, C., M. A. Mausz, and G. Pohnert. 2012. A co-culturing/metabolomics approach to investigate chemically mediated interactions of planktonic organisms reveals influence of bacteria on diatom metabolism. *Metabolomics* **9**: 349–359. doi:10.1007/s11306-012-0453-1
- Prosser, G. a, G. Larrouy-Maumus, and L. P. S. de Carvalho. 2014. Metabolomic strategies for the identification of new enzyme functions and metabolic pathways. *EMBO Rep.* **15**: 657–69. doi:10.15252/embr.201338283
- Rosenwasser, S., M. A. Mausz, D. Schatz, and others. 2014. Rewiring host lipid metabolism by large viruses determines the fate of *Emiliania huxleyi*, a bloom-forming alga in the ocean. *Plant Cell* **26**: 2689–2707. doi:10.1105/tpc.114.125641
- Sañudo-Wilhelmy, S. A., L. S. Cutter, R. Durazo, and others. 2012. Multiple B-vitamin depletion in large areas of the coastal ocean. *Proc. Natl. Acad. Sci. U. S. A.* **109**: 14041–5. doi:10.1073/pnas.1208755109
- Schnoes, A. M., S. D. Brown, I. Dodevski, and P. C. Babbitt. 2009. Annotation error in public databases: Misannotation of molecular function in enzyme superfamilies. *PLoS Comput. Biol.* **5**. doi:10.1371/journal.pcbi.1000605
- Swenson, T. L., S. Jenkins, B. P. Bowen, and T. R. Northen. 2015. Untargeted soil metabolomics methods for analysis of extractable organic matter. *Soil Biol. Biochem.* **80**: 189–198. doi:10.1016/j.soilbio.2014.10.007
- Tweeddale, H., L. Notley-McRobb, and T. Ferenci. 1998. Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool “metabolome” analysis. *J. Bacteriol.* **180**: 5109–5116 PB-. doi:082/4
- Volk, T., and M. I. Hoffert. 1985. Ocean carbon pumps: Analysis of relative strengths and efficiencies in ocean-driven atmospheric CO₂ changes. *Geophys. Monogr. Ser.* **32**: 99–110.
- Wakeham, S. G., J. I. Hedges, C. Lee, M. L. Peterson, and P. J. Hernes. 1997. Compositions and transport of lipid biomarkers through the water column and surficial sediments of the equatorial Pacific Ocean. *Deep. Res. Part II Top. Stud. Oceanogr.* **44**: 2131–2162. doi:10.1016/S0967-0645(97)00035-0
- Wang, X., A. Zhang, and H. Sun. 2013. Power of metabolomics in diagnosis and biomarker discovery of hepatocellular carcinoma. *Hepatology* **57**: 2072–2077. doi:10.1002/hep.26130
- Warne, M. A., E. M. Lenz, D. Osborn, J. M. Weeks, and J. K. Nicholson. 2001. Comparative biochemistry and short-term starvation effects on the earthworms *Eisenia veneta* and *Lumbricus terrestris* studied by ¹H NMR spectroscopy and pattern recognition. *Soil Biol. Biochem.* **33**: 1171–1180. doi:10.1016/S0038-0717(01)00021-9
- Wikoff, W. R., A. T. Anfora, J. Liu, P. G. Schultz, S. A. Lesley, E. C. Peters, and G. Siuzdak. 2009. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc. Natl. Acad. Sci. U. S. A.* **106**: 3698–703. doi:10.1073/pnas.0812874106

Chapter 2.

Extraction efficiency and quantification of dissolved metabolites
in targeted marine metabolomics

Abstract

The field of metabolomics seeks to characterize the suite of small molecules that comprise the end-products of cellular regulation. Recently, metabolomics has expanded beyond biomedical applications to include environmental studies that explore ecological and biogeochemical questions. We have developed a targeted metabolomics method using ESI-LC-MS/MS to analyze metabolites dissolved in seawater. Preparation of samples from the marine environment presents challenges because dilute metabolites must be concentrated and de-salted. We present the extraction efficiencies of 89 metabolites in our targeted method using solid phase extraction (SPE). In addition, we calculate the limits of detection and quantification for the metabolites in the method and compare the instrument response factors in five different matrices ranging from deionized water to spent medium from cultured marine microbes. High background organic matter content reduces the instrument response factor for only a small group of metabolites, yet enhances the extraction efficiency for other metabolites on the SPE cartridge used here, a modified styrene-divinylbenzene polymer called PPL. Aromatic or larger uncharged compounds, in particular, are reproducibly well-retained on the PPL polymer. This method is suitable for the detection of dissolved metabolites in marine samples, with limits of detection ranging from < 1 pM to ~ 2 nM dependent on the dual impacts of seawater matrix on extraction efficiency and on instrument response factors.

2.1 Introduction

Metabolomics is an “omics” technique that seeks to measure the small organic biomolecules produced by cells (Oliver et al. 1998; Fiehn 2002). Because these small molecules are the end-products of multiple levels of metabolic regulation, their concentrations provide a temporal snapshot of the metabolic state or phenotype of an organism. In particular, metabolites produced by non-enzymatic reactions, such as those formed by reaction with a reactive oxygen species, or whose production is regulated by other small molecules, must be monitored directly because their production cannot be inferred from genomic or proteomic information.

Metabolomics can be used as a diagnostic tool, identifying biomarkers of disease within the human metabolome. Examples include discovery of metabolic markers of cancers (Armitage and Barbas 2014) and Crohn’s Disease (Jansson et al. 2009). Metabolomics has also been applied in a wide range of organisms and environments, examining how metabolite abundances respond to environmental factors. In the oceans, marine metabolites have been a valuable source of new natural products, while other metabolomics applications are still rare but growing. For example, recent marine culture experiments have revealed metabolite production not predicted by genomic information (Baran et al. 2010; Fiore et al. 2015), metabolic shifts in response to a specific metabolite (Johnson et al. 2016), and changes in the quantity and composition of metabolite production during co-culturing (Paul et al. 2012). Complementary field studies are now underway in several laboratories to understand microbial activity and organic matter cycling *in situ*. In these studies, metabolomics has great potential to reveal marine microbe phenotypic expression under differing environmental conditions and the chemical interactions by which ecological communities function, as well as the role that these communities play in the marine carbon cycle.

In our laboratory, metabolomics analyses are conducted with complementary untargeted and targeted mass spectrometry techniques (Kido Soule et al. 2015; Longnecker et al. 2015). The targeted method, which is the focus of this study, currently measures 89 metabolites using liquid chromatography tandem mass spectrometry (LC-MS/MS) and an electrospray ionization (ESI) source (Kido Soule et al. 2015). The targeted molecules encompass many classes of metabolites, including amino acids, nucleotides, vitamins, osmolytes, and intermediates of primary metabolism. This paper characterizes the impact of matrix composition (i.e., all of the chemical components of a sample) on both the extraction and analysis of these metabolites.

Metabolomic analysis of seawater samples is challenging due to the complexity of the organic matter, low concentrations of metabolites, and high levels of salt in the matrix, which may affect instrument response through ionization suppression or enhancement. ESI, in particular, is known to be susceptible to such matrix effects. In complex sample matrices, the co-elution of matrix material with the analyte of interest can change the efficiency with which the analyte enters the gas phase relative to the efficiency in pure solvent (King et al. 2000; Taylor 2005; Böttcher et al. 2007). Thus, when using a calibration curve made in pure solvent, the difference in the instrument response factor between the calibration curve and the analyte in the sample matrix can result in inaccurate quantification. For accurate quantification, analytical chemists generally correct for these matrix-specific effects on ionization efficiency by using a standard addition method or a matrix-matched calibration curve with isotopically-labeled internal standards (Stüber and Reemtsma 2004; Kang et al. 2007; Boyd et al. 2008). These methods are effective but not always practical. The standard addition method is not feasible when there is a limited amount of sample material available or when a large number of samples must be analyzed. Similarly, isotopically-labeled internal standards can be prohibitively expensive and

isotopically-labeled metabolites are often not commercially available.

Components of seawater matrices not only affect the mass spectrometer response through ionization suppression or enhancement, but the non-volatile salts can also precipitate, clogging the ESI needle and the mass spectrometer inlet. Working with high salinity samples thus requires an extraction method that will remove salt. Moreover, dissolved metabolites in marine samples are often dilute and require a method that will successfully concentrate molecules with a wide range of physical and chemical properties. Typically, extractions of organic molecules are optimized for a specific compound class such as amino acids or lipids. In these cases the structural similarity of the molecules of interest facilitates selection of an extraction method tailored for that functional group, polarity, or charge. For instance, lipids, which are typically hydrophobic or amphiphilic, can be extracted from water using a non-polar solvent (Cequier-Sánchez et al. 2008), while amino acids can be isolated using a ligand exchange resin (Lee and Bada 1975). However, there are limited options for extracting a structurally diverse array of metabolites using a single technique that can be easily performed in the field.

Solid phase extraction (SPE) has proven to be a straightforward way to extract dissolved organic compounds from seawater with near-complete removal of salt (Lara and Thomas 1994; Dittmar et al. 2008), particularly where large volumes of water must be sampled (we typically extract 4 liters of seawater in the field). In marine organic geochemistry a commonly used SPE substrate is a modified styrene-divinylbenzene polymer called PPL (Agilent Bond Elut PPL). In the marine environment, this polymer has been shown to have a superior extraction efficiency for marine dissolved organic carbon (DOC; 43-62%) compared to other solid phase extraction resins such as C18 and C8 silica-based sorbents (Dittmar et al. 2008). The samples are acidified to pH 2-3 prior to SPE in order to protonate organic acids to improve retention by the sorbent (Dittmar

et al. 2008; Longnecker 2015), although protonation of nitrogen-containing compounds may reduce retention of those molecules. This extraction technique has been used to study the composition of low molecular weight dissolved organic matter (LMW DOM, < 1000 Da) from samples collected throughout the ocean. Thus, using the PPL SPE polymer for dissolved metabolite extractions allows for more direct comparison with previous studies. The PPL polymer is best suited for extraction of uncharged, slightly polar, medium-sized (~100-1000 Da) analytes. Most charged or very small molecules are not well-retained, if at all. However, extraction efficiencies of individual compounds on this polymer cannot be precisely predicted based on molecular structure, highlighting the importance of experimentally determining these parameters.

While PPL SPE is used in a variety of environmental and biomedical applications to extract molecules of interest, studies of extraction efficiency have primarily been confined to a limited set of related molecules without comparison across variable matrix conditions. For example, the extraction efficiencies of acyl homoserine lactones (20-100% recovery) and of drugs such as vancomycin (47% recovery) and furosemide (68% recovery) using PPL SPE have been measured (Li et al. 2006; Baranowska et al. 2010). To our knowledge there has been no comprehensive study of PPL SPE extraction efficiency for a core set of metabolites. Furthermore, there is currently little information available regarding matrix effects on metabolite quantification in marine samples, thus limiting our ability to predict these effects. For researchers applying targeted metabolomics methods in complex matrices, metabolite extraction efficiencies and matrix effects are essential parameters to incorporate into experimental design and interpretation.

Here we measure the extraction efficiency of 89 dissolved metabolites using SPE and

characterize how a range of sample matrices commonly encountered in marine metabolomics affect both extraction efficiency and electrospray ionization efficiency of target metabolites. The five matrices selected range from minimal-organic matter to high organic matter concentrations and from salt-free to the typical salt content of marine samples. The organic matter composition also varies in the matrices from a higher proportion of recalcitrant organic molecules in seawater compared to spent media treatments where small, polar, labile molecules from cell exudates and lysates will dominate. This allows us to examine the matrix effects within common marine sample types, but also to consider the impact of specific matrix parameters on analyte behavior within a wider range of metabolomics samples.

2.2 Materials and methods

2.2.1 Materials

All metabolite standards were obtained from Sigma-Aldrich at the highest purity available with the following exceptions: dimethylsulfonylpropionate (DMSP), which was purchased from Research Plus, Inc.; 2,3-dihydroxypropane-1-sulfonate (DHPS) and acetyltaurine, which were donated by Dr. Mary Ann Moran (University of Georgia); *S*-(1,2-dicarboxyethyl)glutathione, which was purchased from Bachem. All media and artificial seawater components were purchased from Fisher Scientific (ACS Certified) with the exception of sodium orthovanadate from Alexis Biochemicals, ethylenediaminetetraacetic acid, manganese chloride tetrahydrate, zinc sulfate heptahydrate (ReagentPlus, $\geq 99\%$), and cobalt chloride hexahydrate (ACS Reagent) from Sigma-Aldrich, and magnesium sulfate heptahydrate (98%), copper sulfate pentahydrate (ACS Reagent), sodium molybdate dihydrate (99+%), selenious acid (99+%), and potassium chromate (99.5%) from Acros Organics. Hydrochloric acid (trace metal

grade), acetonitrile (Optima grade), and methanol (Optima grade) were purchased from Fisher Scientific. All water was purified by a Milli-Q system (Millipore; resistivity 18.2 M Ω •cm @ 25 °C, TOC < 1 μ M). Glassware was acid washed and combusted in an oven at 460°C for at least 5 h. All plasticware was washed with Citranox and then soaked in a 10% HCl acid bath over night. Plasticware and media stock solutions were autoclaved before use.

2.2.2 Matrices

Five matrices were tested in this study (Table 1). They will be referred to by the acronyms identified here throughout the rest of the paper. A pure water matrix (MQ) was collected from the Milli-Q system. The artificial seawater matrix (Turks Island Salts; TIS) was constituted in Milli-Q water as follows (per liter of water): 28 g sodium chloride, 670 mg potassium chloride, 5.5 g magnesium chloride hexahydrate, 6.9 g magnesium sulfate heptahydrate, 1.45 g calcium chloride dehydrate. The seawater matrix (Vineyard Sound Seawater; VSW) was collected from Vineyard Sound, MA, U.S.A. and filtered through a 0.2 μ m filter (Omnipore (PTFE), EMD Millipore). The heterotrophic bacterium spent medium matrix (Rpom) consisted of spent medium from a culture of *Ruegeria pomeroyi* sacrificed during the stationary phase of growth and filtered through a 0.2 μ m filter (Omnipore, EMD Millipore). Similarly, the autotrophic spent medium matrix (Mp) consisted of spent medium from a culture of *Micromonas pusilla* sacrificed during stationary growth phase and filtered through a 0.1 μ m filter (Omnipore, EMD Millipore). See Table S1 for media recipe details.

2.2.3 Sample preparation for extraction efficiency determination

Triplicate bottles (polycarbonate for MQ, TIS, and VSW matrices and glass jars for Rpom and Mp matrices) of each matrix were spiked with a standard mix (~50:50 water:methanol) of metabolites of interest (see experimental design in Figure 1 and Table S2 for

volumes and concentrations) and acidified to pH 2-3 with 12 M hydrochloric acid (Dittmar et al. 2008; Longnecker 2015). The spike concentrations were in the nM range to ensure determination of extraction efficiency although many metabolites are measured at pM concentrations in the ocean. An additional triplicate set of bottles without the standard metabolite mix spike was used as a control. The bottles were shaken and allowed to sit for 30 min. Each solution was then loaded onto a methanol-rinsed 1 g Agilent Bond Elut PPL cartridge (6 mL cartridge volume) and the water was pulled through the cartridge using a vacuum pump. The cartridge was rinsed with 24 mL 0.01 M HCl, and the metabolites were eluted with 6 mL methanol (Dittmar et al. 2008). All samples for LC-MS/MS analysis (control and standard mix extracts) were dried down in a vacufuge and reconstituted in 95:5 water:acetonitrile (see Table S2 for volumes; 1% of the final volume was 5 µg/mL biotin-d₂ as an injection standard). An aliquot of the control extract (without biotin-d₂) was used to make matrix-matched calibration curves for each matrix treatment and to determine the bulk organic carbon extraction efficiency of the PPL cartridge (see below; Figure 1). The matrix-matched calibration curves contained 9 calibration points ranging from 0.5 ng/mL to 1000 ng/mL. Further discussion of how the calibration curves were made can be found in Table S3, Text S1, and Figure S1.

2.2.4 Determination of limits of detection

Data from 9 calibration curves made in Milli-Q water collected over 6 months (April-September 2015) were used to calculate a limit of detection (LOD). This is defined as (Boyd et al. 2008):

$$LOD = stdA \times tvalue$$

where stdA is the standard deviation of the concentration of the analyte at a selected low concentration and tvalue refers to the Student's t-test interval for a 1-tailed t-test with $\alpha = 0.01$

and n-1 degrees of freedom. For the majority of metabolites, the 1 ng/mL calibration point was used to calculate the standard deviation. However, in some cases a higher concentration had to be used. This method requires that the concentration used to calculate the LOD be 1-5 times higher than the calculated LOD. The LOD was calculated at multiple low concentrations for each metabolite to determine what concentration was acceptable to use and would provide an accurate LOD (Table S4). Outliers were identified using a Generalized Extreme Studentized Deviate Test and up to two outliers were excluded from the calculation. The limit of quantification (LOQ) was calculated with respect to the LOD by using the definition (Boyd et al. 2008):

$$LOQ = \frac{10}{3} \times LOD$$

The LOD and LOQ in the other matrices could not be calculated directly due to the presence of analytes in the matrix itself. Instead, we used the ratio of the analyte signal in a given matrix to the analyte signal in MQ to adjust the LOD and LOQ calculated in MQ, according to the following calculation:

$$eLOD_i \text{ or } eLOQ_i = \frac{LOD_{mq} \text{ or } LOQ_{mq}}{\frac{mA_i}{mA_{mq}}}$$

where $eLOD_i$ is the estimated limit of detection in matrix i, $eLOQ_i$ is the estimated limit of quantification in matrix i, LOD_{mq} is the LOD calculated in MQ, mA_i is the mean of the quadruplicate response of the analyte at 5 ng/mL (or 50 ng/mL if no peak intensity at 5 ng/mL) in matrix i with the matrix blank subtracted, and mA_{mq} is the mean of the quadruplicate response of the analyte in MQ at the same concentration. This type of ratio has been proposed to evaluate matrix effects (Rogatsky and Stein 2005) although it has not been used to estimate LOD values to our knowledge. $eLOD_i$ and $eLOQ_i$ values were only calculated for the VSW and Rpom matrices.

2.2.5 Organic carbon measurements

The total organic carbon concentrations of the samples were measured before and after the PPL extraction. Prior to extraction, 40 mL water samples of each matrix type (in triplicate) were acidified to pH 2-3 using 12 M hydrochloric acid. To determine the amount of organic carbon retained on the PPL cartridge, the remaining PPL-extracted control sample (not spiked with metabolite standard mix, as described above) was dried down in the vacufuge and brought up in 1 mL of Milli-Q water. 400 μ L of that 1 mL was added to a vial of 25 mL Milli-Q water and 25 μ L of 12 M hydrochloric acid. Samples were stored at 4°C until analysis. All samples were analyzed using a Shimadzu TOC-V_{CSH} Total Organic Carbon Analyzer, according to standard practices. A five-point calibration curve made with potassium hydrogen phthalate was used and blanks were run regularly. Duplicate injections had an average coefficient of variability of < 1%. Comparisons to standards from D. Hansell (University of Miami) were made daily.

2.2.6 Mass spectrometry

LC-MS/MS analysis was performed on a Phenomenex C18 reversed phase column (Synergi Fusion, 2.1 x 150 mm, 4 μ m) coupled via a heated ESI source to a triple quadrupole mass spectrometer (Thermo Scientific TSQ Vantage) operated under selected reaction monitoring mode (SRM) with polarity switching (Kido Soule et al. 2015). Quantification and confirmation SRM transitions were monitored for each metabolite. Eluent A was Milli-Q water with 0.1% (v/v) formic acid and Eluent B was acetonitrile with 0.1% (v/v) formic acid. The following gradient was used: hold at 5% B for 2 minutes; ramp to 65% B for 16 minutes; ramp to 100% B for 7 minutes and hold for 8 min. An 8.5 minute column re-equilibration with the starting ratio of eluents was carried out between sample analyses. Between the different matrices a Milli-Q water blank and three samples of unspiked matrix extract were run to rinse and

condition the column. Composite chromatograms of a sample from each matrix type are shown as examples in the supplemental information (Figure S2).

2.2.7 Data processing

XCalibur RAW files from the mass spectrometer were converted to mzML files using MSConvert (Chambers et al. 2012). MAVEN (Melamud et al. 2010; Clasquin et al. 2012) was used to select and integrate peaks. Peaks below a MAVEN quality threshold of 0.4 (on a scale of 0-1) were discarded. To enhance confidence in metabolite identification, quantification and confirmation peaks were required to have retention times within 12 seconds (0.2 minutes) of each other. Future data processing will also include a peak quality control check for the confirmation ion. Calibration curves were required to have at least five calibration points and points were selected to go from the lowest concentration level to one concentration level above the highest concentration detected in an experimental sample (Table S5). To compare the response factors (slopes) between matrices, the full calibration curve up to 1000 ng/mL was used. Spiked metabolite concentrations were calculated directly from calibration curves without subtracting metabolite concentrations in matrix controls because the calibration curves were matrix-matched and so already incorporated any necessary correction for the background concentration of the metabolite. Extraction efficiencies of individual metabolites were calculated using the following formula:

$$\%EE = \frac{CF_i}{CT_i} \times 100$$

where %EE is the percent extraction efficiency, CF_i is the final concentration of the analyte measured in ng/mL, and CT_i was the target concentration of the analyte assuming 100% retention (either 500 ng/mL or 1500 ng/mL, see Table S2 and Table S7).

2.2.8 Structural characterization

To link the extraction efficiency of analytes to their structure, their structural characteristics were assessed in a number of ways. SPARC (<http://archemcalc.com/>), an online computational tool, was used to calculate the charge and partitioning coefficients of metabolites based on their structure. Specifically, these calculations were carried out at pH 2, as this was at the extreme end of the pH range used during SPE, and accounted for the ionic strength of seawater. K_{oc} (the partitioning coefficient between water and generic organic carbon) was calculated for metabolites that were well-retained on the PPL cartridges to examine whether metabolites that had improved extraction efficiencies in the culture matrices had similar water-organic carbon partitioning coefficients. The partitioning coefficient for hexadecane and water was also calculated for metabolites to determine if it had a relationship with the measured extraction efficiencies. This data is not included as no relationship was found. Other information such as polar surface area and molecular weight were obtained from the online chemical database ChemSpider; but yielded no relationship with extraction efficiency for our metabolites.

2.3 Assessment

2.3.1 Matrix characterization

The five matrices used in this study were chosen to examine how organic matter concentration and type, as well as salt, influence analyte behavior throughout the extraction and analysis processes. In particular, we compared minimal-organic matter and salt-free matrices to common matrices encountered in marine metabolomics samples. These include field samples from the ocean and spent seawater media from controlled laboratory experiments. We measured the organic carbon content of each matrix before and after SPE to have a quantitative measure of the carbon-based differences between the selected matrices (Table 1). The MQ treatment

provides a salt- and organic matter-free (1 μM) baseline comparison. Similarly, in order to separately examine the effect of salt, TIS contains minimal organic matter (13 μM) but has the same salt concentration as seawater. The rest of the matrix treatments were a range of sample types containing both salt and organic matter. The VSW matrix had a dissolved organic carbon content of 107 μM , and a large portion is recalcitrant and relatively non-polar (compared to the culture matrices) due to its low heteroatom (oxygen and nitrogen) content (reviewed by Carlson and Hansell 2015; Repeta 2015).

The organic carbon measurements before and after PPL extraction were used to determine a bulk dissolved organic carbon (DOC) extraction efficiency for each of the matrices. The DOC extraction efficiency in the VSW matrix was higher (51%) than either of the spent media matrix samples (Table 1). In contrast, the bulk DOC extraction efficiency in the Rpom matrix was poor (5%). The high initial DOC concentration (1865 μM) is due to the addition of propionate (initial carbon concentration from propionate in the media was 3000 μM) as an organic carbon substrate to fuel the growth of *R. pomeroyi*. Since this molecule is small and highly polar, it is poorly retained by PPL SPE. The Mp matrix did not contain an added carbon substrate and so the initial DOC pool (597 μM) is primarily composed of cellular exudates and lysates, which explains the higher extraction efficiency (24%). Due to the variable bulk extraction efficiencies of organic carbon, the relative differences in organic carbon concentration were muted in the extracts relative to the initial matrices (final organic carbon contents in the 6 mL (1 mL for spent media) extracts: 9 mM (VSW), 23 mM (Rpom), 35 mM (Mp)).

2.3.2 Instrument response factors

The impact of these matrices on the ionization efficiency of each metabolite in our targeted metabolomics method can be characterized by comparing the instrument response factor

(i.e. the slope of the calibration curve) in each matrix to that in MQ (Figure 2; Table S6). To compare the linear regressions of the analyte calibration curves in TIS, VSW, Rpom, and Mp with those of the analytes in MQ, we used a wild bootstrap method (Estévez-Pérez et al. 2016), which generated the p -values reported in Table S6. Calibration curve slopes for an analyte were considered significantly different between a matrix and MQ when p -values were less than 0.05 (Figure 2). Estévez-Pérez et al. (2016) found this random resampling approach to be the most consistent way to compare linear calibration curves in cases such as this, where there are relatively few calibration points and the data are not homoscedastic.

In general, more metabolites exhibit ionization suppression in the spent media matrices (Rpom and Mp) compared to TIS and VSW (Figure 2). This appears linked to the total DOC content of the matrices as the Rpom and Mp extracts contained 23 and 35 mM DOC, respectively (excluding the spiked metabolites), compared to the lower DOC concentrations of 0.4 and 9 mM in the TIS and VSW extracts. However, it is possible that differences in organic matter composition between the matrices also play a role. When the distribution of instrument response factor ratios in each matrix is compared, there is a significant difference between the distributions in the two media matrices compared to those of the non-culture matrices (Figure S3). Therefore, as we would predict, increased matrix DOC content affects the ionization efficiency of some metabolites; but this impact is not statistically significant for most metabolites on an individual basis. Even in the most extreme case (Mp), 73% of metabolites were unaffected.

There are 10, 4, 11, and 21 metabolites in the TIS, VSW, Rpom, and Mp matrices, respectively, that have response factors that are lower than in MQ and 2, 6, 3, and 3 metabolites with response factors that are higher than in MQ, all with p -values below 0.05 (Figure 2). The affected metabolites elute from the column at a wide range of retention times and vary between

the matrices, making it difficult to predict which metabolites will be affected. Some metabolites are relatively consistent across matrices; putrescine, for instance, which elutes from the column early, has a significantly elevated response factor in three of the four matrices, while cyanocobalamin, with a relatively late retention time, has a significantly decreased response factor in all four matrices. Taurocholic acid, with a retention time of 17 minutes, is only significantly reduced in the Mp matrix, while malic acid, with a retention time of 1.9 minutes, has decreased ionization efficiencies in both Rpom and Mp. These are notable observations because not only do these metabolites have different polarities and functional groups, but they are also ionized in distinct matrix environments due to their different retention times. These results suggest that changes in response factor are linked to DOC content but are also molecule-specific. However, for the purpose of method implementation, when the majority of metabolites are considered, the effect of the matrix on instrument response at the measured DOC concentrations is not significant and thus does not require a matrix-matched calibration approach. Rather, standard curves in Milli-Q water provide good quantification (within a factor of 2) for most metabolites in this study.

2.3.3 Extraction efficiency

Extraction efficiencies for targeted metabolites by PPL SPE were determined in the five matrices (Table S7). As shown in Figure 3, there is a general consensus across all matrices between LC retention time and extraction efficiency; metabolites with later retention times are better retained on the PPL cartridge. For example, in the VSW and Rpom matrices, consistent SPE retention of all analytes is not reached until ~2.5 minutes in the chromatogram (Figure 3). However, the extraction efficiencies of a few metabolites run contrary to this general trend. The extraction efficiency for *S*-(5'-adenosyl)-L-homocysteine (SAH) is greater than 10% but its

column retention time is less than 2.5 minutes. Conversely, a small group of metabolites have column retention times greater than 2.5 minutes but have extraction efficiencies less than 10%: xanthine, inosine 5'-monophosphate (IMP), inosine, guanosine, *N*-acetylglutamic acid, and desthiobiotin.

The structural attributes that appear to distinguish metabolites that are retained on the PPL resin from those that are not seem to be a combination of aromaticity, charge at pH 2, and secondarily molecular weight (Figure 3; Table S8). Overall, aromatic compounds or those that are neutral at pH 2 seem to be best retained. For example, SAH's predominant species is positively charged (a small fraction is amphiphilic) at pH 2, but it is an aromatic compound and retained by the PPL resin. However, other aromatic compounds are not well-retained by PPL, and size and charge may be factors in these cases. For example, xanthine is quite small, as it is a purine nucleobase and lacks a sugar functional group. In contrast, inosine and guanosine contain a sugar moiety but their positive charge may offset the size benefit of the sugar. IMP contains a phosphate group, which may make it too polar to be retained, despite its aromaticity. Other metabolites, such as *N*-acetylglutamic acid and desthiobiotin, are neutral at pH 2, but are small and not aromatic, perhaps leading to their poor retention. A partitioning coefficient that shows a relationship with the measured extraction efficiencies has not been identified. For example, the partitioning coefficient (log D) for hexadecane and water does not explain the extraction efficiencies. While there are structural differences that appear to govern metabolite retention on the SPE polymer, the variable impacts of aromaticity, charge, and molecular size preclude straightforward prediction of retention dynamics at this time.

In addition to metabolite chemical and physical properties, the sample matrix can affect metabolite extraction efficiency (Figure 4). In particular, of the 35 metabolites that were retained

on the PPL cartridge, 14 had higher extraction efficiencies in the culture matrices (Rpom and Mp; see Table S7) than in the other three matrices (MQ, TIS, VSW). For instance, thymidine and xanthosine both show improved extraction efficiency in the culture matrices (Figure 4B). We hypothesize that colloidal organic matter in the culture matrices could be trapped on the polymer and provide additional sites for the metabolites to interact. Although we could determine a structural trend for metabolites that are best retained on the PPL, similar probing of structural patterns of metabolites that showed improved extraction efficiency in the culture matrices did not yield any clear relationship. Characteristics such as aromaticity, charge, molecular weight, log K_{oc} (pH 2), and polar surface area of metabolites were examined (Table S9) without uncovering a clear relationship.

The extraction efficiencies reported here will serve as a useful guide to researchers intending to extract metabolites using PPL SPE from samples with similar organic carbon content. To our knowledge this study is the most extensive look at the extraction efficiencies for a core set of metabolites using SPE. However, the variability in extraction efficiency across matrix types indicates that care must be taken when extrapolating these values to other sample matrices.

2.3.4 Limits of detection and quantification

The instrument limit of detection (LOD) provides a conservative estimate of the concentration at which a metabolite can be confidently detected, but not reliably quantified. In contrast, the instrument limit of quantification (LOQ) is defined to be a higher (more conservative) level where we expect reliable quantification. LOD and LOQ were calculated for all metabolites in Milli-Q water using the EPA definitions (Table 2; Glaser et al. 1981; EPA 1997; Boyd et al. 2008). This definition of the LOD ensures with 99% confidence that the

analyte is greater than zero; however, there is a 50% chance of a false negative because any measurement that falls on the low side of a normal distribution will be below the limit. LOD values range from 0.2 ng/mL for caffeine to 48 ng/mL for nicotinamide adenine dinucleotide phosphate (NADP). As a general rule metabolites with early retention times have higher LODs due to increased baseline noise in the early part of the chromatogram. Polyamines (putrescine, ornithine) and metabolites with phosphate groups in particular tend to have low quality peaks that are difficult to integrate at low concentrations.

2.4 Discussion

This study examined analyte behavior at a number of points in the process of sample preparation and analysis for a targeted metabolomics method. Ultimately, the efficiency of these processes limits the concentrations of metabolites we can measure in the ocean and in cultures. Due to the presence of the analytes in the non-MQ matrices, the LOD and LOQ could not be calculated directly for those matrices, but were instead estimated (eLOD: Table 2; eLOQ: Table S4) to account for ion enhancement or suppression in each matrix. The eLODs for many metabolites in the VSW and Rpom matrices are similar to the calculated LOD in MQ, however there seem to be more metabolites with elevated eLODs and greater uncertainty in the Rpom matrix compared to the VSW matrix (Table 2).

In Table 2, the lowest concentrations we can expect to measure in the ocean for a given metabolite using our sampling method are reported, based on the adjusted eLOD for VSW. Both the total concentration that could be measured assuming 100% extraction efficiency and the concentration that can be measured accounting for the PPL SPE extraction efficiency in VSW are calculated. While this paper has focused on extraction of dissolved metabolites, we have

previously determined ~100% extraction efficiency for particulate or intracellular metabolites (Kido Soule et al. 2015). For the dissolved metabolites, concentrations range from <1 pM to ~2400 pM (or ~2.4 nM). The highest detection limits are constrained by extremely low SPE extraction efficiencies. Of course, if the focus is an individual metabolite that is poorly retained on PPL, another extraction method should be found. Nonetheless these values suggest that our method can measure low concentrations of many metabolites in the field, an essential capability in the marine environment. For example, riboflavin, a B vitamin, has been measured at concentrations of 0.5-7 pM in the ocean (Sañudo-Wilhelmy et al. 2012). Accounting for riboflavin's extraction efficiency and eLOD in VSW, our detection limit in the ocean is 0.6 pM, allowing us to adequately detect typical concentrations in the ocean. Concentrations of phenylalanine around 1.5 nM have been measured in estuarine environments (Coffin 1989), which is well above the calculated detection limit of 15 pM for our method. This targeted metabolomics method will allow us to undertake the task of mapping the oceanic distributions and concentrations of these molecules, many of which have never been measured in the ocean.

2.5 Conclusions

The impact of solution matrix composition on the extraction efficiency and instrument response factor of the analytes was unexpected. The extraction efficiency differed significantly depending on the organic matter content of the matrix, with 40% of the metabolites retained by the polymer having enhanced extraction efficiencies in the culture matrices compared to MQ, TIS, and VSW. Higher matrix DOC concentrations were also linked to increased ionization suppression and enhancement relative to MQ, with 11% of metabolites having a significantly different response factor in VSW compared to 27% in Mp. This trend might be exacerbated at

even higher matrix DOC concentrations than those studied here.

This dataset provides practical information that can be used to back-calculate estimates of initial analyte concentrations and also characterizes the impact of a number of representative marine matrices. For those working in marine or aquatic systems, the PPL SPE extraction efficiency trends observed should be particularly useful, as they will allow researchers to anticipate whether an analyte is likely to be observed using this method, and also further informs our understanding of the structural selectivity of this extraction method which may be relevant to past and future studies that characterize dissolved organic matter extracted with this polymer. While the gold standard for absolute metabolite quantification within a complex matrix remains the standard addition method, these results suggest that, in the cases analyzed here, the matrix does not compromise quantification.

Acknowledgments

This work was funded by the Gordon and Betty Moore Foundation (Grant #3304 to EBK) and the National Science Foundation (Grant OCE-1154320 to EBK and Krista Longnecker). Funding is also gratefully acknowledged from Simons Foundation International. The instruments in the WHOI FT-MS Facility were purchased with support from the GBMF and NSF. Support for WMJ came from a National Defense Science and Engineering Fellowship. The authors would also like to thank the following people for their assistance with experimental design, lab work, and data analysis and interpretation: Krista Longnecker, Cara Fiore, Yina Liu, David Griffith, and Philip Gschwend.

References

- Armitage, E. G., and C. Barbas. 2014. Metabolomics in cancer biomarker discovery: current trends and future perspectives. *J. Pharm. Biomed. Anal.* **87**: 1–11. doi:10.1016/j.jpba.2013.08.041
- Baran, R., B. P. Bowen, N. J. Bouskill, E. L. Brodie, S. M. Yannone, and T. R. Northen. 2010. Metabolite identification in *Synechococcus* sp. PCC 7002 using untargeted stable isotope assisted metabolite profiling. *Anal. Chem.* **82**: 9034–9042. doi:10.1021/ac1020112
- Baranowska, I., A. Wilczek, and J. Baranowski. 2010. Rapid UHPLC method for simultaneous determination of vancomycin, terbinafine, spironolactone, furosemide and their metabolites: application to human plasma and urine. *Anal. Sci.* **26**: 755–9. doi:10.2116/analsci.26.755
- Böttcher, C., E. Roepenack-Lahaye, E. Willscher, D. Scheel, and S. Clemens. 2007. Evaluation of matrix effects in metabolite profiling based on capillary liquid chromatography electrospray ionization quadrupole time-of-flight mass spectrometry. *Anal. Chem.* **79**: 1507–1513.
- Boyd, R. K., C. Basic, and R. A. Bethem. 2008. Trace quantitative analysis by mass spectrometry, John Wiley & Sons Ltd.
- Carlson, C. A., and D. A. Hansell. 2015. DOM sources, sinks, reactivity, and budgets, p. 65–102. *In* D.A. Hansell and C.A. Carlson [eds.], *Biogeochemistry of marine dissolved organic matter*. Elsevier.
- Cequier-Sánchez, E., C. Rodríguez, Á. G. Ravelo, and R. Zárate. 2008. Dichloromethane as a solvent for lipid extraction and assessment of lipid classes and fatty acids from samples of different natures. *J. Agric. Food Chem.* **56**: 4297–4303.
- Chambers, M. C., B. Maclean, R. Burke, and others. 2012. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**: 918–920. doi:10.1038/nbt.2377
- Clasquin, M. F., E. Melamud, and J. D. Rabinowitz. 2012. LC-MS data processing with MAVEN: A metabolomic analysis and visualization engine. *Curr. Protoc. Bioinforma.* **37**: 1–23. doi:10.1002/0471250953.bi1411s37
- Coffin, R. B. 1989. Bacterial uptake of dissolved free and combined amino acids in estuarine waters. *Limnol. Oceanogr.* **34**: 531–542.
- Dittmar, T., B. Koch, N. Hertkorn, and G. Kattner. 2008. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol. Oceanogr. Methods* **6**: 230–235. doi:10.4319/lom.2008.6.230
- EPA. 1997. Guidelines establishing test procedures for the analysis of pollutants (app. B, Part 136, Definition and procedures for determination of the method detection limit): U.S. Code of Federal Regulations, Title 40.
- Estévez-Pérez, G., J. M. Andrade, and R. R. Wilcox. 2016. Bootstrap approach to compare the slopes of two calibrations when few standards are available. *Anal. Chem.* **88**: 2289–2295. doi:10.1021/acs.analchem.5b04004
- Fiehn, O. 2002. Metabolomics—the link between genotypes and phenotypes. *Plant Mol. Biol.* **48**: 155–71.
- Fiore, C. L., K. Longnecker, M. C. Kido Soule, and E. B. Kujawinski. 2015. Release of ecologically relevant metabolites by the cyanobacterium, *Synechococcus elongatus* CCMP 1631. *Environ. Microbiol.* **17**: 3949–3963. doi:10.1111/1462-2920.12899
- Glaser, J. A., D. L. Foerst, G. D. McKee, S. A. Quave, and W. L. Budde. 1981. Trace analyses for wastewaters. *Environ. Sci. Technol.* **15**: 1426–1435.

- Jansson, J., B. Willing, M. Lucio, A. Fekete, J. Dicksved, J. Halfvarson, C. Tysk, and P. Schmitt-Kopplin. 2009. Metabolomics reveals metabolic biomarkers of Crohn's disease. *PLoS One* **4**: e6386. doi:10.1371/journal.pone.0006386
- Johnson, W. M., M. C. Kido Soule, and E. B. Kujawinski. 2016. Evidence for quorum sensing and differential metabolite production by the marine heterotroph, *Ruegeria pomeroyi*, in response to DMSP. *ISME J.* **10**: 2304–2316. doi:10.1038/ismej.2016.6
- Kang, J., L. A. Hick, and W. E. Price. 2007. Using calibration approaches to compensate for remaining matrix effects in quantitative liquid chromatography/electrospray ionization multistage mass spectrometric analysis of phytoestrogens in aqueous environmental samples. *Rapid Commun. Mass Spectrom.* **21**: 4065–4072. doi:10.1002/rcm
- Kido Soule, M. C., K. Longnecker, W. M. Johnson, and E. B. Kujawinski. 2015. Environmental metabolomics: analytical strategies. *Mar. Chem.* **177**: 374–387. doi:10.1016/j.marchem.2015.06.029
- King, R., R. Bonfiglio, C. Fernandez-Metzler, C. Miller-Stein, and T. Olah. 2000. Mechanistic investigation of ionization suppression in electrospray ionization. *J. Am. Soc. Mass Spectrom.* **11**: 942–950. doi:10.1016/S1044-0305(00)00163-X
- Lara, R. J., and D. N. Thomas. 1994. Isolation of marine dissolved organic matter: Evaluation of sequential combinations of XAD resins 2, 4, and 7. *Anal. Chem.* **66**: 2417–2419.
- Lee, C., and J. L. Bada. 1975. Amino acids in equatorial Pacific Ocean water. *Earth Planet. Sci. Lett.* **26**: 61–68. doi:10.1016/0012-821X(75)90177-6
- Li, X., A. Fekete, M. Englmann, and others. 2006. Development and application of a method for the analysis of N-acylhomoserine lactones by solid-phase extraction and ultra high pressure liquid chromatography. *J. Chromatogr. A* **1134**: 186–193. doi:10.1016/j.chroma.2006.09.047
- Longnecker, K. 2015. Dissolved organic matter in newly formed sea ice and surface seawater. *Geochim. Cosmochim. Acta* **171**: 39–49. doi:10.1016/j.gca.2015.08.014
- Longnecker, K., J. Futrelle, E. Coburn, M. C. Kido Soule, and E. B. Kujawinski. 2015. Environmental metabolomics: Databases and tools for data analysis. *Mar. Chem.* **177**: 366–373. doi:10.1016/j.marchem.2015.06.012
- Melamud, E., L. Vastag, and J. D. Rabinowitz. 2010. Metabolomic analysis and visualization engine for LC-MS data. *Anal. Chem.* **82**: 9818–9826.
- Oliver, S. G., M. K. Winson, D. B. Kell, and F. Baganz. 1998. Systematic functional analysis of the yeast genome. *Trends Biotechnol.* **16**: 373–378.
- Paul, C., M. A. Mausz, and G. Pohnert. 2012. A co-culturing/metabolomics approach to investigate chemically mediated interactions of planktonic organisms reveals influence of bacteria on diatom metabolism. *Metabolomics* **9**: 349–359. doi:10.1007/s11306-012-0453-1
- Repeta, D. J. 2015. Chemical characterization and cycling of dissolved organic matter, p. 41–47. *In* D.A. Hansell and C.A. Carlson [eds.], *Biogeochemistry of marine dissolved organic matter*. Elsevier.
- Rogatsky, E., and D. Stein. 2005. Evaluation of matrix effect and chromatography efficiency: New parameters for validation of method development. *J. Am. Soc. Mass Spectrom.* **16**: 1757–1759. doi:10.1016/j.jasms.2005.07.012
- Sañudo-Wilhelmy, S. A., L. S. Cutter, R. Durazo, and others. 2012. Multiple B-vitamin depletion in large areas of the coastal ocean. *Proc. Natl. Acad. Sci. U. S. A.* **109**: 14041–5. doi:10.1073/pnas.1208755109
- Stüber, M., and T. Reemtsma. 2004. Evaluation of three calibration methods to compensate

matrix effects in environmental analysis with LC-ESI-MS. *Anal. Bioanal. Chem.* **378**: 910–916. doi:10.1007/s00216-003-2442-8

Taylor, P. J. 2005. Matrix effects: The Achilles heel of quantitative high-performance liquid chromatography-electrospray-tandem mass spectrometry. *Clin. Biochem.* **38**: 328–334. doi:10.1016/j.clinbiochem.2004.11.007

Table 1. Experimental treatments and their dissolved organic carbon (DOC) concentrations before and after PPL solid phase extraction (SPE) extraction

| Treatment ID | MQ | TIS | VSW | Rpom | Mp |
|------------------------------------|-----------|--------------------|-------------------------|-------------------------------------------|--------------------------------------------|
| Matrix | Milli-Q | Turks Island Salts | Vineyard Sound Seawater | Spent media from <i>Ruegeria pomeroyi</i> | Spent media from <i>Micromonas pusilla</i> |
| DOC (μ M) Before Extraction | 1 | 13 | 107 | 1865 | 597 |
| Volume Extracted (mL) | 1000 | 1000 | 1000 | 250 | 250 |
| DOC (μ M) in Extract | 200 | 400 | 9000 | 23000 | 35000 |
| DOC Bulk Extraction Efficiency (%) | --- | 20 | 51 | 5 | 24 |
| Extract Volume (mL) | 6 | 6 | 6 | 1 | 1 |

Table 2. Instrumental limits of detection (LOD) and limits of quantification (LOQ) in MQ, estimated (eLOD) for the VSW and Rpom matrices,^a and limits of detection for field samples.^{b,c} (--- indicates cases where the LOD could not be calculated due to high variability in the data and n.r. indicates metabolites that are not retained on the PPL polymer)

| Compound | LOD (ng/mL) | LOQ (ng/mL) | VSW eLOD (ng/mL) | Rpom eLOD (ng/mL) | LOD in Seawater 100% EE (pM) ^b | LOD in Seawater (from PPL, pM) ^c |
|----------------------------------|----------------|----------------|------------------------|-------------------------|----------------------------------------------------|------------------------------------------------------|
| (6R)-5,6,7,8-tetrahydrobiopterin | --- | --- | --- | --- | --- | --- |
| 1-deoxy-D-xylulose-5-phosphate | 7.4 | 25 | 12±2 | 27±18 | 6.9±1.2 | n.r. |
| 2,3-dihydroxybenzoic acid | 8.3 | 28 | 6.4±0.4 | 6.8±0.5 | 5.2±0.4 | 33±3 |
| 2,3-dihydroxypropane-1-sulfonate | 5.9 | 20 | 29±72 | 36±147 | 23±58 | n.r. |
| 3-mercaptopropionic acid | 11 | 38 | 12±2 | 10±3 | 12±2 | 78±15 |
| 4-aminobenzoic acid | 0.63 | 2.1 | 0.65±0.03 | 0.69±0.04 | 0.59±0.03 | 7.9±1.9 |
| 4-hydroxybenzoic acid | 17 | 58 | 16±3 | 18±4 | 15±2 | 95±16 |
| 5-methylthioadenosine | 0.80 | 2.7 | 0.67±0.04 | 0.59±0.18 | 0.28±0.02 | 2.0±0.4 |
| 6-phosphogluconic acid | 44 | 150 | 48±42 | 114±141 | 22±19 | n.r. |
| acetyltaurine | 7.5 | 25 | 9±3 | 26±10 | 6.6±2.0 | n.r. |
| adenine | 1.9 | 6.3 | 2.2±0.6 | 7±9 | 2.0±0.6 | n.r. |
| adenosine | 1.5 | 5.0 | 1.4±0.2 | 2.0±0.4 | 0.68±0.11 | 120±23 |
| adenosine 5'-monophosphate | 2.4 | 7.9 | 17±34 | 10.3±0.9 | 6±12 | n.r. |
| alpha-ketoglutaric acid | 5.8 | 19 | 39±194 | 16±13 | 33±166 | n.r. |
| arginine | 2.3 | 7.7 | 7±4 | 7±5 | 5±3 | n.r. |
| aspartic acid | --- | --- | --- | --- | --- | --- |
| betaine | 2.9 | 9.7 | 3.9±0.7 | 4.2±0.5 | 4.2±0.7 | n.r. |
| biotin | 2.3 | 7.5 | 1.5±0.2 | 1.5±0.2 | 0.76±0.11 | 8.0±1.3 |
| caffeine | 0.18 | 0.58 | 0.15±0.03 | 0.15±0.03 | 0.097±0.016 | 2.5±0.5 |
| chitobiose | 5.7 | 19 | 5±6 | 11±16 | 1.5±1.8 | n.r. |
| chitotriose | 20 | 66 | 13±3 | 49±40 | 2.5±0.5 | n.r. |
| choline | 1.9 | 6.2 | 1.7±0.2 | 2.2±0.4 | 2.1±0.3 | n.r. |
| ciliatine | 2.5 | 8.2 | 3.3±0.9 | 3.4±0.6 | 3.3±0.9 | n.r. |
| citric acid | 41 | 140 | 24±9 | 22±6 | 15±6 | n.r. |
| citrulline | 2.0 | 6.5 | 2.7±0.8 | 6±2 | 1.9±0.6 | n.r. |
| cyanocobalamin | 2.1 | 6.9 | 1.2±0.3 | 1.3±0.6 | 0.11±0.02 | 0.8±0.2 |
| cysteine | 8.5 | 28 | 11.4±1.8 | 29±10 | 11.7±1.8 | n.r. |
| cytosine | 2.3 | 7.6 | 1.9±1.4 | 2.7±1.8 | 2.2±1.5 | n.r. |
| D-(-)3-phosphoglyceric acid | 24 | 81 | 10±11 | 58±125 | 7±7 | n.r. |
| D-glucosamine 6-phosphate | 3.7 | 12 | 5.1±1.6 | 0.58±0.05 | 2.5±0.8 | n.r. |
| D-ribose 5-phosphate | 5.2 | 17 | 7.1±9.8 | 12±13 | 4±5 | n.r. |
| desthiobiotin | 0.75 | 2.5 | 0.56±0.04 | 0.2±0.5 | 0.33±0.03 | 25±16 |
| dihydroxyacetone phosphate | 4.9 | 16 | 1±3 | 6±12 | 1.1±1.8 | n.r. |
| DMSP ⁴ | 0.99 | 3.3 | 1.16±0.10 | 1.95±0.17 | 1.08±0.09 | n.r. |
| ectoine | 0.85 | 2.8 | 0.75±0.09 | 0.71±0.10 | 0.66±0.08 | n.r. |
| folic acid | 0.57 | 1.9 | 0.35±0.08 | 0.33±0.08 | 0.10±0.02 | 0.7±0.3 |
| fosfomycin | 27 | 89 | 59±19 | --- | 53±17 | n.r. |
| fumaric acid | --- | --- | --- | --- | --- | --- |
| GABA ³ | 1.4 | 4.8 | 1.3±0.3 | 2.0±0.4 | 1.5±0.3 | n.r. |
| glucose 6-phosphate | 8.9 | 30 | 12.9±1.7 | 64±87 | 6.2±0.8 | n.r. |

| Compound | LOD (ng/mL) | LOQ (ng/mL) | VSW eLOD (ng/mL) | Rpom eLOD (ng/mL) | LOD in Seawater 100% EE (pM) ^b | LOD in Seawater (from PPL, pM) ^c |
|-------------------------------------------|----------------|----------------|------------------------|-------------------------|----------------------------------------------------|------------------------------------------------------|
| glutamic acid | 36 | 120 | 19±10 | 45±36 | 17±9 | n.r. |
| glutamine | 0.86 | 2.9 | 1.29±0.18 | 1.6±0.4 | 1.10±0.16 | n.r. |
| glutathione | 5.1 | 17 | 7±9 | 80±32 | 3±4 | n.r. |
| glutathione oxidized | 15 | 50 | 5±6 | 42±98 | 1.0±1.3 | n.r. |
| glyphosate | 5.1 | 17 | 3±5 | 15±26 | 2±3 | n.r. |
| guanine | 2.1 | 6.9 | 1.8±0.3 | 14±10 | 1.5±0.3 | n.r. |
| guanosine | 0.44 | 1.5 | 0.39±0.10 | 0.46±0.12 | 0.17±0.05 | 17±5 |
| indole 3-acetic acid | 1.1 | 3.7 | 1.07±0.11 | 1.06±0.11 | 0.77±0.08 | 7.0±0.8 |
| inosine | 2.0 | 6.8 | 2.4±0.7 | 3.3±1.6 | 1.1±0.3 | 76±30 |
| inosine 5'-monophosphate | 9.4 | 31 | 5.7±1.0 | 8±2 | 1.8±0.3 | n.r. |
| (iso)leucine | 4.0 | 13 | 4±3 | 2.5±1.7 | 4±3 | n.r. |
| kynurenine | 1.6 | 5.3 | 1.3±0.2 | 1.60±0.18 | 0.79±0.13 | 8.3±1.9 |
| malic acid | 26 | 85 | 27±22 | 30±22 | 26±21 | n.r. |
| methionine | 4.8 | 16 | 48±17 | 58±25 | 40±14 | n.r. |
| muramic acid | --- | --- | --- | --- | --- | --- |
| <i>N</i> -acetylglucosamine | 16 | 54 | 16±4 | 76±62 | 9±2 | n.r. |
| <i>N</i> -acetylglutamic acid | 5.5 | 18 | 6±15 | 7±4 | 4±10 | 2400±6000 |
| <i>N</i> -acetylmuramic acid | 21 | 70 | 18±4 | 17±4 | 7.8±1.9 | 1500±1300 |
| NAD ² | --- | --- | --- | --- | --- | --- |
| NADP ¹ | 49 | 160 | 16±19 | 28±33 | 3±3 | n.r. |
| ornithine | 3.1 | 10 | 6±4 | 12±19 | 5±4 | n.r. |
| orotic acid | 30 | 100 | 27±8 | 53±19 | 22±7 | n.r. |
| pantothenic acid | 15 | 51 | 11.8±1.0 | 11.6±1.0 | 6.7±0.6 | 80±15 |
| phenylalanine | 0.67 | 2.2 | 0.73±0.11 | 0.3±0.2 | 0.56±0.09 | 15±3 |
| phosphoenolpyruvate | --- | --- | --- | --- | --- | --- |
| proline | 2.1 | 6.9 | 2.0±0.5 | 3.7±1.2 | 2.2±0.5 | n.r. |
| putrescine | 31 | 100 | 39±7 | 46±14 | 55±10 | n.r. |
| pyridoxine | 1.4 | 4.5 | 1.4±0.3 | 1.18±0.20 | 1.1±0.2 | 236±45 |
| riboflavin | 0.63 | 2.1 | 0.35±0.04 | 0.3±0.2 | 0.115±0.014 | 0.65±0.08 |
| <i>S</i> -(1,2-dicarboxyethyl)glutathione | 4.1 | 14 | 1.4±1.9 | 2±3 | 0.4±0.6 | n.r. |
| <i>S</i> -(5'-adenosyl)-L-homocysteine | 1.9 | 6.3 | 4±4 | 18±68 | 1.4±1.2 | 72±64 |
| <i>S</i> -adenosyl methionine | 7.6 | 25 | 6.2±1.9 | 5.0±1.7 | 1.9±0.6 | n.r. |
| sarcosine | 23 | 75 | 34±13 | 42±15 | 48±19 | n.r. |
| serine | 1.8 | 6.0 | 2.9±0.9 | 3.6±1.8 | 3.5±1.0 | n.r. |
| <i>sn</i> -glycerol 3-phosphate | 9.6 | 32 | 9±8 | 21±22 | 7±6 | n.r. |
| succinic acid | --- | --- | --- | --- | --- | --- |
| taurine | 3.4 | 11 | 0.4±0.4 | 0.5±0.5 | 0.4±0.4 | n.r. |
| taurocholic acid | 5.3 | 18 | 2.3±0.8 | 2.5±0.8 | 0.57±0.20 | 5±2 |
| thiamin | 39 | 130 | 36±5 | 19±3 | 17±2 | n.r. |
| thiamin monophosphate | 21 | 69 | 18±3 | 18±3 | 6.1±1.0 | n.r. |
| threonine | 1.7 | 5.8 | 2.5±1.2 | 2.8±1.0 | 2.6±1.2 | n.r. |
| thymidine | 19 | 64 | 20±7 | 21±6 | 10±3 | 123±52 |
| tryptamine | 0.43 | 1.4 | 0.42±0.03 | 0.45±0.03 | 0.32±0.02 | 2.7±0.3 |
| tryptophan | 5.3 | 18 | 6.1±0.8 | 5.5±0.4 | 3.7±0.5 | 30±5 |
| tyrosine | 3.2 | 11 | 3.3±1.4 | 8±18 | 2.3±0.9 | 1140±550 |
| uracil | 4.4 | 15 | 5±4 | --- | 6±4 | n.r. |
| uridine 5'-monophosphate | 40 | 130 | --- | --- | --- | --- |
| xanthine | 2.0 | 6.8 | 1.8±0.7 | 2.2±0.4 | 1.5±0.5 | n.r. |
| xanthosine | 2.1 | 6.9 | 2.0±0.4 | 2.0±0.7 | 0.88±0.17 | 47±11 |

^a LOD and LOQ were calculated in Milli-Q water as described in the Experimental Section. The estimated

(eLOD) values for the VSW and Rpom matrices were adjusted from the LOD values as described in the Experimental Section. ^bThese values were calculated to reflect the LOD (using the VSW eLOD) in the initial seawater sample assuming that 100% of the analyte was extracted from 4 L of water and that the final extract was 500 μ L. ^cThis LOD represents the lowest value we would expect to detect in a field sample given the PPL SPE extraction efficiencies determined here and our current sampling protocol. 4 L of seawater is extracted, then eluted with 6 mL of methanol, dried down, and reconstituted in 3 mL 95:5 water:acetonitrile. This was calculated only for metabolites with at least a 1% extraction efficiency.

¹Nicotinamide adenine dinucleotide phosphate. ²Nicotinamide adenine dinucleotide. ³ γ -aminobutyric acid.

⁴Dimethylsulfoniopropionate

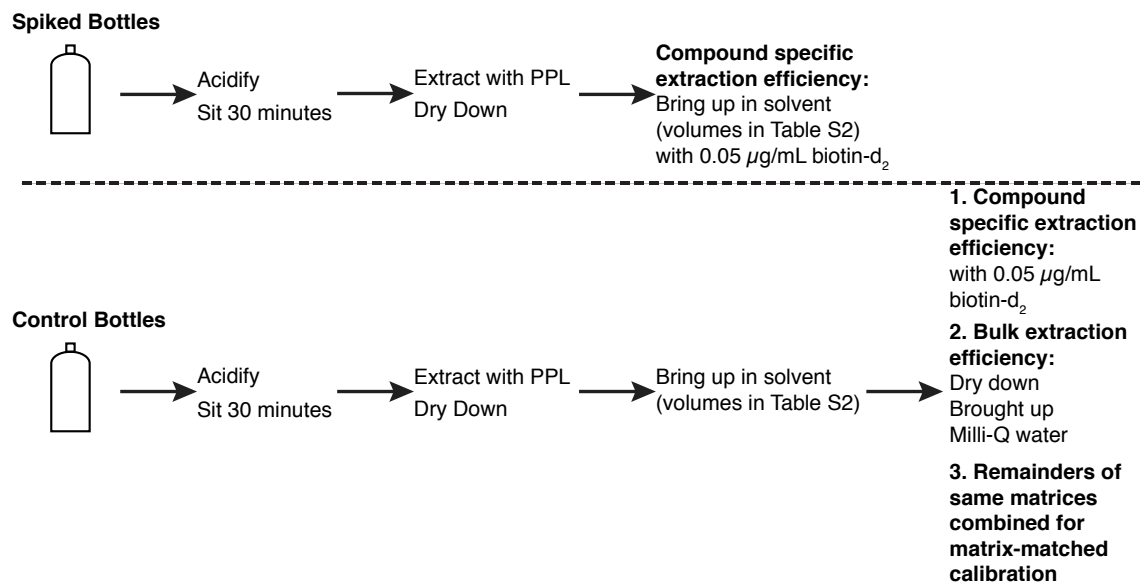


Figure 1. Experimental design.

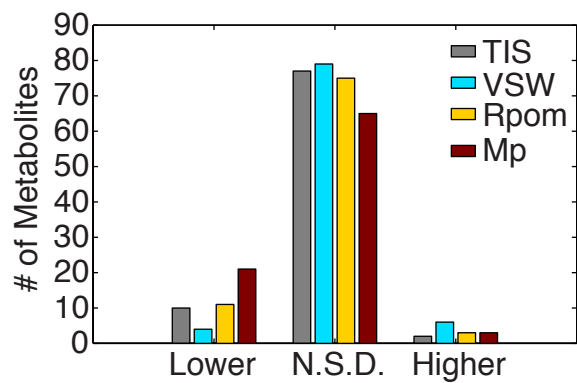


Figure 2. The number of metabolites with a lower, higher, or not significantly different (N.S.D.) instrument response factor in each matrix relative to the MQ matrix. The p -value was considered significant if $p < 0.05$.

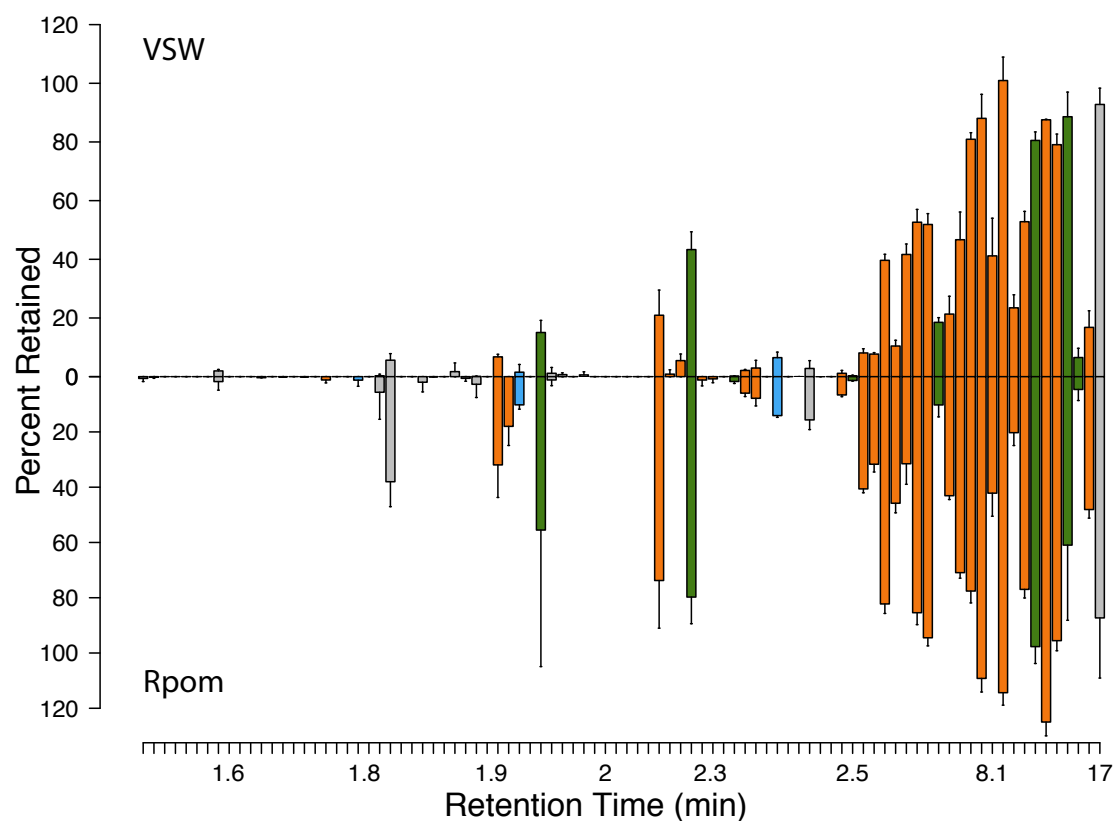


Figure 3. The extraction efficiency for each metabolite plotted by retention time (not scaled) for the VSW and Rpom matrices. Orange bars indicate compounds with aromatic functional groups, green bars are compounds that are neutrally charged at pH 2 but not aromatic, blue are non-aromatic compounds whose charge could not be determined, and gray bars are compounds that are charged at pH 2 or have an extremely polar functional group like phosphate.

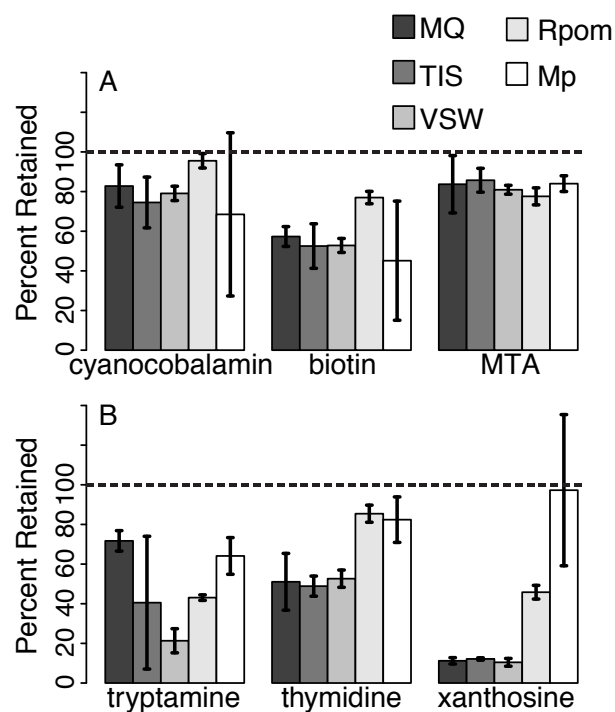


Figure 4. Comparison of extraction efficiencies of select metabolites across all five matrices. (A) Examples of metabolites that are consistently retained in all matrices. MTA: 5-methylthioadenosine. (B) Examples of metabolites that are variably retained in different matrices (tryptamine) or have enhanced extraction efficiencies in the spent media matrices (thymidine and xanthosine).

Supplemental Information

Text S1. Methods for making calibration curves.

When the matrix-matched calibration curves were made using the volumes shown in Table S3, the concentration of extract was not consistent throughout the calibration curve. Working solutions were made at concentrations ranging from 1000 ng/mL to 5 ng/mL, where the solvent was a mixture of extract and the original water:methanol-based stock solutions. Working solutions were further diluted with extract to the final metabolite concentration. This resulted in the highest calibration points containing less (more dilute) extract than the lowest calibration points. For instance, the highest calibration point (1000 ng/mL) was 50% more dilute in extract (i.e. sample matrix) than the lowest calibration point. Intermediate calibration points contained increasing sample matrix with decreasing calibration level (Table S3). We compared the calibration curves where the matrix concentration (run 4 months later on the same instrument) was the same throughout the curve to the above curves, in order to ensure that this inconsistency did not significantly change the calibration curves. We found no significant difference according to an ANOVA, when comparing the variability of the ratios of the response factors of the Milli-Q water calibration curves to those in VSW and Rpom (Figure S1). This suggests that the difference in calibration curve protocol did not change the response.

Table S1. Recipes for growth media for *Ruegeria pomeroyi* and *Micromonas pusilla* (made in 0.2 µm-filtered and autoclaved seawater).

| <i>Ruegeria pomeroyi</i> Medium | | <i>Micromonas pusilla</i> Medium | |
|----------------------------------|------------------------|------------------------------------------------------|-------------------------|
| Component | Concentration | Component | Concentration |
| Na propionate | 1.0×10^{-3} M | NaNO ₃ | 8.82×10^{-4} M |
| NH ₄ Cl | 4.0×10^{-3} M | NaH ₂ PO ₄ · H ₂ O | 3.62×10^{-5} M |
| NaH ₂ PO ₄ | 3.0×10^{-8} M | Na ₂ EDTA · 2H ₂ O | 1.17×10^{-5} M |
| FeCl ₂ -EDTA | 1×10^{-7} M | FeCl ₃ · 6H ₂ O | 1.17×10^{-5} M |
| ZnCl ₂ | 1×10^{-7} M | MnCl ₂ · 4 H ₂ O | 9.09×10^{-7} M |
| MnCl ₂ | 1×10^{-7} M | ZnSO ₄ · 7H ₂ O | 8.00×10^{-8} M |
| CoCl ₂ | 1×10^{-9} M | CoCl ₂ · 6H ₂ O | 5.00×10^{-8} M |
| f/2 vitamin solution | 1 mL / L medium | CuSO ₄ · 5H ₂ O | 1.00×10^{-8} M |
| | | Na ₂ MoO ₄ · 2H ₂ O | 8.22×10^{-8} M |
| | | H ₂ SeO ₃ | 1.00×10^{-8} M |
| | | NiSO ₄ · 6H ₂ O | 1.00×10^{-8} M |
| | | Na ₃ VO ₄ | 1.00×10^{-8} M |
| | | K ₂ CrO ₄ | 1.00×10^{-8} M |
| | | f/2 vitamin solution | 1 mL / L medium |

Table S2. Metabolite spike parameters in extraction efficiency experiment. *Volumes were selected to best reflect the volumes representative of different types of sampling (VSW = field sampling, Rpom and Mp = laboratory cultures). See Scheme 1 for treatment of spiked bottles and control bottles.

| | MQ | TIS | VSW | Rpom | Mp |
|--------------------------------------------------------------|-----------|------------|------------|-------------|-----------|
| Bottle Volumes (L) | 1 | 1 | 1 | 0.2 | 0.2 |
| Metabolite concentration (ng/mL) | 3 | 3 | 3 | 2.5 | 2.5 |
| Elevated* metabolite concentration (ng/mL) | 9 | 9 | 9 | 7.5 | 7.5 |
| Final Extract Volume (mL) | 6 | 6 | 6 | 1 | 1 |
| Target metabolite concentration in extract (ng/mL) | 500 | 500 | 500 | 500 | 500 |
| Elevated* target metabolite concentration in extract (ng/mL) | 1500 | 1500 | 1500 | 1500 | 1500 |

*Higher (3X) concentrations of metabolites were used for those compounds that we hypothesized would be retained on the PPL cartridge only at very low concentrations. Metabolites marked with an asterisk in Table S7 were spiked at the elevated concentration.

Table S3. Volumes and components used to make calibration curves. Post-PPL extract (i.e., sample matrix) was used to reconstitute these solutions (see Scheme 1). There were 5 initial 10 µg/mL solutions in ~50:50 water:methanol. The internal standard was biotin-d₂. The working solutions were used to make the final calibration samples and were diluted with extract. The final concentrations were also made with additional extract. “Final Ratio of Extract” indicates the final proportion of the total sample that was post-PPL extract.

Working Solutions:

| WS # | Final Concentration (ng/mL) | Solution Used | Volume of Solution (µL) | Ratio Extract | Volume of Extract (µL) | Total Volume (µL) |
|-------------|------------------------------------|----------------------|--------------------------------|----------------------|-------------------------------|--------------------------|
| WS1 | 1000 | 10 µg/ml mixes | 50 of each mix | 0.5 | 250 | 500 |
| WS2 | 500 | WS1 | 100 | 0.75 | 100 | 200 |
| WS3 | 100 | WS1 | 25 | 0.95 | 225 | 250 |
| WS4 | 50 | WS2 | 25 | 0.97 | 225 | 250 |
| WS5 | 10 | WS3 | 25 | 0.99 | 225 | 250 |
| WS6 | 5 | WS4 | 25 | 1.00 | 225 | 250 |

Calibration Solutions:

| Final Concentration (ng/mL) | Solution Used | Volume of Solution (µL) | Final Ratio of Extract | Volume of Extract (µL) | Volume of Internal Standard (µL) | Total Volume (µL) |
|------------------------------------|----------------------|--------------------------------|-------------------------------|-------------------------------|-----------------------------------------|--------------------------|
| 1000 | 10 µg/ml mixes | 100 | 0.49 | 98 | 2 | 200 |
| 500 | WS1 | 100 | 0.74 | 98 | 2 | 200 |
| 250 | WS2 | 100 | 0.86 | 98 | 2 | 200 |
| 100 | WS1 | 20 | 0.94 | 178 | 2 | 200 |
| 50 | WS2 | 20 | 0.96 | 178 | 2 | 200 |
| 10 | WS3 | 20 | 0.98 | 178 | 2 | 200 |
| 5 | WS4 | 20 | 0.99 | 178 | 2 | 200 |
| 1 | WS5 | 20 | 0.99 | 178 | 2 | 200 |
| 0.5 | WS6 | 20 | 0.99 | 178 | 2 | 200 |

Table S4. Estimates of instrumental limits of quantification (eLOQ) in VSW and Rpom and the concentration used to calculate the Milli-Q water limits of detection (LOD). This is additional information for Table 2. Column 3 is the calibration concentration used to determine the LOD in Milli-Q water. (--- indicate cases where the LOD could not be calculated due to high variability in the data)

| Compound | VSW eLOQ (ng/mL) | Rpom eLOQ (ng/mL) | ng/mL |
|----------------------------------|------------------|-------------------|-------|
| (6R)-5,6,7,8-tetrahydrobiopterin | --- | --- | --- |
| 1-deoxy-D-xylulose-5-phosphate | 40±7 | 90±59 | 10 |
| 2,3-dihydroxybenzoic acid | 21.4±1.4 | 22.7±1.5 | 50 |
| 2,3-dihydroxypropane-1-sulfonate | 96±239 | 119±491 | 10 |
| 3-mercaptopropionic acid | 39±7 | 32±10 | 50 |
| 4-aminobenzoic acid | 2.17±0.11 | 2.30±0.14 | 5 |
| 4-hydroxybenzoic acid | 54±9 | 59±13 | 50 |
| 5-methylthioadenosine | 2.22±0.13 | 2.0±0.6 | 1 |
| 6-phosphogluconic acid | 160±140 | 380±469 | 50 |
| acetyltaurine | 29±9 | 87±34 | 10 |
| adenine | 7±2 | 24±29 | 5 |
| adenosine | 4.8±0.8 | 6.6±1.2 | 5 |
| adenosine 5'-monophosphate | 56±113 | 34±3 | 10 |
| alpha-ketoglutaric acid | 130±648 | 55±44 | 5 |
| arginine | 22±12 | 25±17 | 5 |
| aspartic acid | --- | --- | --- |
| betaine | 13±2 | 14.0±1.8 | 5 |
| biotin | 5.0±0.7 | 4.9±0.6 | 5 |
| caffeine | 0.50±0.08 | 0.50±0.08 | 1 |
| chitobiose | 17±20 | 37±52 | 5 |
| chitotriose | 43±9 | 162±132 | 50 |
| choline | 5.7±0.7 | 7.4±1.4 | 5 |
| ciliatine | 11±3 | 11.4±1.9 | 5 |
| citric acid | 79±31 | 73±20 | 50 |
| citrulline | 9±3 | 20±7 | 5 |
| cyanocobalamin | 4.0±0.9 | 4±2 | 5 |
| cysteine | 38±6 | 96±34 | 10 |
| cytosine | 6±5 | 9±6 | 5 |
| D-(-)3-phosphoglyceric acid | 33±35 | 192±416 | 50 |
| D-glucosamine 6-phosphate | 17±5 | 1.92±0.16 | 5 |
| D-ribose 5-phosphate | 24±33 | 40±42 | 5 |
| desthiobiotin | 1.87±0.14 | 0.6±1.7 | 1 |
| dihydroxyacetone phosphate | 5±8 | 19±42 | 10 |
| DMSP ⁴ | 3.9±0.3 | 6.5±0.6 | 1 |
| ectoine | 2.5±0.3 | 2.4±0.3 | 1 |

| | | | |
|-------------------------------------------|-----------|---------|-----|
| folic acid | 1.2±0.3 | 1.1±0.3 | 1 |
| fosfomycin | 195±64 | --- | 50 |
| fumaric acid | --- | --- | --- |
| GABA ³ | 4.2±0.9 | 6.5±1.4 | 5 |
| glucose 6-phosphate | 43±6 | 214±291 | 10 |
| glutamic acid | 66±34 | 151±120 | 50 |
| glutamine | 4.3±0.6 | 5.4±1.4 | 1 |
| glutathione | 25±31 | 265±106 | 5 |
| glutathione oxidized | 16±21 | 140±326 | 50 |
| glyphosate | 11±16 | 49±87 | 10 |
| guanine | 5.9±1.0 | 48±32 | 5 |
| guanosine | 1.3±0.3 | 1.5±0.4 | 1 |
| indole 3-acetic acid | 3.6±0.4 | 3.5±0.4 | 5 |
| inosine | 8±2 | 11±5 | 5 |
| inosine 5'-monophosphate | 19±3 | 26±7 | 10 |
| (iso)leucine | 14±9 | 8±6 | 5 |
| kynurenine | 4.3±0.7 | 5.3±0.6 | 5 |
| malic acid | 91±74 | 100±73 | 50 |
| methionine | 160±57 | 194±84 | 5 |
| muramic acid | --- | --- | --- |
| <i>N</i> -acetylglucosamine | 54±12 | 253±206 | 50 |
| <i>N</i> -acetylglutamic acid | 20±51 | 22±13 | 5 |
| <i>N</i> -acetylmuramic acid | 61±15 | 55±12 | 50 |
| NAD ² | --- | --- | --- |
| NADP ¹ | 53±63 | 93±108 | 50 |
| ornithine | 19±14 | 38±64 | 5 |
| orotic acid | 91±27 | 177±62 | 50 |
| pantothenic acid | 39±3 | 39±3 | 50 |
| phenylalanine | 2.0±0.4 | 1.1±0.7 | 1 |
| phosphoenolpyruvate | --- | --- | --- |
| proline | 6.8±1.6 | 12±4 | 5 |
| putrescine | 130±22 | 155±46 | 50 |
| pyridoxine | 4.8±0.9 | 3.9±0.7 | 5 |
| riboflavin | 1.16±0.14 | 1.1±0.7 | 1 |
| <i>S</i> -(1,2-dicarboxyethyl)glutathione | 5±6 | 7±11 | 5 |
| <i>S</i> -(5'-adenosyl)-L-homocysteine | 14±12 | 59±226 | 5 |
| <i>S</i> -adenosyl methionine | 21±6 | 17±6 | 10 |
| sarcosine | 113±44 | 139±50 | 50 |
| serine | 10±3 | 12±6 | 5 |
| <i>sn</i> -glycerol 3-phosphate | 32±26 | 70±73 | 10 |
| succinic acid | --- | --- | --- |

| | | | |
|--------------------------|-----------|-----------|----|
| taurine | 1.2±1.2 | 1.7±1.7 | 5 |
| taurocholic acid | 8±3 | 8±3 | 5 |
| thiamin | 120±16 | 63±9 | 50 |
| thiamin monophosphate | 62±11 | 59±11 | 50 |
| threonine | 8±4 | 9±3 | 5 |
| thymidine | 68±22 | 70±20 | 50 |
| tryptamine | 1.38±0.08 | 1.49±0.11 | 1 |
| tryptophan | 20±3 | 18.5±1.3 | 5 |
| tyrosine | 11±5 | 25±60 | 5 |
| uracil | 17±12 | --- | 5 |
| uridine 5'-monophosphate | --- | --- | 50 |
| xanthine | 6±2 | 7.3±1.4 | 5 |
| xanthosine | 6.7±1.3 | 7±2 | 5 |

¹Nicotinamide adenine dinucleotide phosphate. ²Nicotinamide adenine dinucleotide. ³ γ -aminobutyric acid.

⁴Dimethylsulfoniopropionate.

Table S5. The r^2 values and the number of points used to calculate each calibration curve that was used for the extraction efficiency concentrations. MQ, TIS, VSW, Rpom, Mp columns show the calibration curve information for each matrix type.

| Compound | MQ | | TIS | | VSW | | Rpom | | Mp | |
|----------------------------------|------------------|---|------------------|---|------------------|----|------------------|----------------|------------------|----------------|
| | Cal. Curve r^2 | n | Cal. Curve r^2 | n | Cal. Curve r^2 | n | Cal. Curve r^2 | n | Cal. Curve r^2 | n |
| (6R)-5,6,7,8-tetrahydrobiopterin | 0.9851 | 6 | 0.9907 | 5 | 0.9975 | 6 | 0.9844 | 6 | 0.9496 | 5 |
| 1-deoxy-D-xylulose-5-phosphate | 0.9996 | 5 | 0.9925 | 5 | 0.9682 | 6 | 0.9762 | 6 | 0.9999 | 5 |
| 2,3-dihydroxybenzoic acid | 0.9744 | 8 | 0.9898 | 9 | 0.9997 | 10 | 0.9587 | 9 | 0.9995 | 6 |
| 2,3-dihydroxypropane-1-sulfonate | 0.9852 | 6 | 0.9773 | 6 | 0.9869 | 6 | 0.2706* | 5 | 0.6870* | 4 [§] |
| 3-mercaptopropionic acid | 0.9962 | 8 | 0.8919 | 7 | 0.9986 | 9 | 0.9977 | 8 | 0.9976 | 8 |
| 4-aminobenzoic acid | 0.9999 | 9 | 0.9992 | 9 | 0.9995 | 7 | 0.9973 | 7 | 1 | 6 |
| 4-hydroxybenzoic acid | 0.9995 | 7 | 0.9996 | 7 | 0.9994 | 9 | 0.9996 | 10 | 0.9974 | 7 |
| 5-methylthioadenosine | 0.9971 | 9 | 0.9987 | 9 | 0.9847 | 9 | 0.9911 | 9 | 0.9871 | 9 |
| 6-phosphogluconic acid | 0.9986 | 6 | 0.9687 | 5 | 0.9976 | 6 | 0.9943 | 6 | 0.9823 | 5 |
| acetyltaurine | 0.9885 | 6 | 0.9773 | 6 | 0.9988 | 6 | 0.9914 | 5 | 0.995 | 5 |
| adenine | 0.9968 | 6 | 0.9789 | 5 | 0.9779 | 6 | 0.9907 | 6 | 0.8770* | 6 |
| adenosine | 0.9787 | 6 | 0.9588 | 8 | 0.9555 | 8 | 0.9932 | 4 [§] | 0.9434 | 7 |
| adenosine 5'-monophosphate | 0.9981 | 6 | 0.9995 | 6 | 0.9999 | 5 | 0.9948 | 6 | 0.9103 | 6 |
| alpha-ketoglutaric acid | 0.9662 | 6 | 0.992 | 5 | 0.9651 | 5 | 0.9985 | 6 | 0.8555* | 5 |
| arginine | 0.9999 | 5 | 0.9936 | 6 | 0.998 | 5 | 0.9948 | 5 | 0.7398* | 7 |
| aspartic acid | 0.9929 | 6 | 0.9441 | 5 | 0.9105 | 5 | 0.9933 | 6 | 0.9703 | 6 |
| betaine | 0.9982 | 6 | 0.9974 | 6 | 0.9976 | 6 | 0.9983 | 6 | 0.9996 | 6 |
| biotin | 0.9986 | 8 | 0.9956 | 9 | 0.9998 | 9 | 0.9493 | 7 | 0.9995 | 8 |
| caffeine | 0.9999 | 8 | 0.9993 | 8 | 0.9997 | 8 | 0.9994 | 8 | 1 | 8 |
| chitobiose | 0.9853 | 6 | 0.9486 | 5 | 0.906 | 5 | 0.9589 | 6 | 0.8722* | 6 |
| chitotriose | 0.9955 | 6 | 0.9666 | 6 | 0.9865 | 6 | 0.9487 | 6 | 0.9053 | 6 |
| choline | 0.9833 | 6 | 0.9803 | 6 | 0.993 | 6 | 0.9992 | 6 | 0.9986 | 6 |
| ciliatine | 0.9916 | 6 | 0.9926 | 6 | 0.9922 | 6 | 0.9877 | 6 | 0.9986 | 6 |
| citric acid | 0.9992 | 6 | 0.8994 | 6 | 0.9353 | 6 | 0.9872 | 6 | 0.9883 | 6 |
| citrulline | 0.9983 | 6 | 0.9997 | 6 | 0.9961 | 6 | 0.9953 | 6 | 0.9988 | 6 |
| cyanocobalamin | 0.9988 | 9 | 0.9992 | 9 | 0.9997 | 9 | 0.9896 | 10 | 0.9992 | 8 |
| cysteine | 0.9984 | 6 | 0.9834 | 5 | 0.9955 | 6 | 0.9873 | 6 | 0.9926 | 6 |
| cytosine | 0.9985 | 6 | 0.9997 | 6 | 0.9996 | 6 | 0.9984 | 6 | 0.9993 | 6 |
| D-(-)3-phosphoglyceric acid | 0.9869 | 5 | 0.9532 | 5 | 0.9846 | 5 | 0.9007 | 6 | 0.9404 | 3 [§] |
| D-glucosamine 6-phosphate | 0.9977 | 6 | 0.9809 | 5 | 0.9986 | 5 | 0.969 | 6 | 0.9827 | 6 |
| D-ribose 5-phosphate | 0.9962 | 6 | 0.9985 | 5 | 0.9994 | 6 | 0.979 | 6 | 0.991 | 6 |

| | | | | | | | | | | |
|-----------------------------------|--------|----------------|--------|----------------|-------------|----|--------|----|---------|----------------|
| desthiobiotin | 0.9988 | 7 | 0.9998 | 6 | 0.9995 | 7 | 0.9957 | 6 | 0.9983 | 8 |
| dihydroxyacetone phosphate | 0.9897 | 6 | 0.9588 | 6 | 0.9738 | 6 | 0.9999 | 5 | 0.9997 | 4 [§] |
| DMSF ⁴ | 0.9999 | 6 | 0.9993 | 6 | 0.9978 | 6 | 0.9993 | 6 | 0.982 | 6 |
| ectoine | 0.9978 | 6 | 0.996 | 6 | 0.9985 | 6 | 0.9999 | 6 | 0.9994 | 6 |
| folic acid | 0.998 | 10 | 0.9977 | 9 | 0.9999 | 9 | 0.9989 | 9 | 0.9998 | 9 |
| fosfomycin | 0.9902 | 4 [§] | 0.9415 | 4 [§] | 0.9931 | 6 | 0.9921 | 5 | 0.9292 | 6 |
| fumaric acid | 0.9862 | 5 | 0.9881 | 5 | 0.7660 * | 3 | 0.9745 | 5 | 0.9434 | 6 |
| GABA ³ | 0.9853 | 6 | 0.9515 | 6 | 0.9917 | 6 | 0.9822 | 6 | 0.9994 | 6 |
| glucose 6-phosphate | 0.9988 | 5 | 0.9978 | 6 | 0.9792 | 6 | 0.986 | 6 | 0.9251 | 5 |
| glutamic acid | 0.9977 | 6 | 0.9983 | 6 | 0.9967 | 6 | 0.9925 | 5 | 0.9977 | 6 |
| glutamine | 0.9884 | 6 | 0.9988 | 6 | 0.9981 | 6 | 0.9984 | 6 | 0.9875 | 5 |
| glutathione | 0.997 | 6 | 0.9769 | 6 | 0.9576 | 6 | 0.9749 | 6 | 0.6877* | 6 |
| glutathione oxidized | 0.9892 | 6 | 0.9913 | 5 | 0.9969 | 6 | 0.9901 | 6 | 0.8029* | 6 |
| glyphosate | 0.9948 | 6 | 0.9803 | 6 | 0.975 | 6 | 0.952 | 5 | 0.8834* | 6 |
| guanine | 0.9994 | 6 | 0.9952 | 6 | 0.9996 | 6 | 0.9864 | 6 | 0.9994 | 7 |
| guanosine | 0.9998 | 6 | 0.9994 | 6 | 0.9995 | 6 | 0.9984 | 8 | 0.9952 | 7 |
| indole 3-acetic acid | 0.9999 | 9 | 0.9311 | 8 | 0.9999 | 8 | 0.9999 | 9 | 0.9498 | 7 |
| inosine | 0.9982 | 6 | 0.9984 | 7 | 0.9976 | 7 | 0.9966 | 9 | 0.9935 | 8 |
| inosine 5'-monophosphate | 0.9982 | 6 | 0.9841 | 6 | 0.9961 | 6 | 0.9936 | 6 | 0.9621 | 6 |
| (iso)leucine | 0.9935 | 6 | 0.9889 | 6 | 0.9838 | 6 | 0.9949 | 5 | 0.9732 | 6 |
| kynurenine | 0.9988 | 9 | 0.9997 | 9 | 0.9995 | 7 | 0.9998 | 7 | 0.9983 | 6 |
| malic acid | 0.9259 | 6 | 0.998 | 6 | 0.9896 | 5 | 0.9988 | 5 | 0.9835 | 6 |
| methionine | 0.9994 | 6 | 0.9969 | 6 | 0.9975 | 6 | 0.9855 | 6 | 0.3817* | 5 |
| muramic acid | 0.9945 | 6 | 0.985 | 5 | 0.998 | 6 | 0.9856 | 5 | 0.9706 | 6 |
| N-acetylglucosamine | 0.9965 | 6 | 0.9849 | 6 | 0.9612 | 6 | 0.9758 | 6 | 0.9733 | 6 |
| N-acetylglutamic acid | 0.9345 | 6 | 0.9993 | 6 | 0.999 | 6 | 0.9977 | 6 | 0.9905 | 6 |
| N-acetylmuramic acid | 0.9988 | 6 | 0.9998 | 6 | 0.9947 | 6 | 0.9983 | 6 | 0.985 | 6 |
| NAD ² | 0.9981 | 6 | 0.9995 | 6 | 0.9398 | 6 | 0.9574 | 6 | 0.9564 | 7 |
| NADP ¹ | 0.9951 | 5 | 0.9976 | 5 | 0.9992 | 6 | 0.9646 | 5 | 0.9597 | 5 |
| ornithine | 0.9955 | 6 | 0.9957 | 6 | 0.9971 | 6 | 0.961 | 5 | 0.9983 | 5 |
| orotic acid | 0.9832 | 6 | 0.9708 | 6 | 0.9932 | 6 | 0.9617 | 6 | 0.9999 | 5 |
| pantothenic acid | 1 | 7 | 0.9997 | 7 | 0.9999 | 8 | 0.9994 | 7 | 0.9985 | 9 |
| phenylalanine | 0.9984 | 8 | 0.9997 | 9 | 0.9996 | 8 | 0.9986 | 8 | ---* | 9 |
| phosphoenolpyruvate | 0.9943 | 5 | 0.9442 | 6 | 0.9765 | 6 | 0.9906 | 3 | 0.9977 | 4 [§] |
| proline | 0.9995 | 6 | 0.998 | 6 | 0.9981 | 6 | 0.9984 | 5 | 0.9994 | 5 |
| putrescine | 0.9977 | 6 | 0.9932 | 6 | 0.9787 | 6 | 0.9944 | 6 | 0.9874 | 6 |
| pyridoxine | 0.9987 | 6 | 0.987 | 8 | 0.9976 | 8 | 0.9317 | 10 | 0.9831 | 10 |
| riboflavin | 0.9988 | 10 | 0.9973 | 9 | 0.9981 | 9 | 0.9866 | 9 | 0.9988 | 9 |
| S-(1,2-dicarboxyethyl)glutathione | 0.9988 | 6 | 0.9322 | 6 | 0.9937 | 6 | 0.9965 | 6 | 0.9757 | 4 [§] |
| S-(5'-adenosyl)-L- | 0.9986 | 6 | 0.9984 | 9 | 0.9908 | 10 | 0.9568 | 6 | 0.9476 | 9 |

| | | | | | | | | | | |
|--------------------------|---------|---|--------|----------------|--------|----|--------|----------------|---------|----------------|
| homocysteine | | | | | | | | | | |
| S-adenosyl methionine | 0.9384 | 6 | 0.9981 | 5 | 0.9529 | 6 | 0.9981 | 5 | 0.9945 | 6 |
| sarcosine | 0.9866 | 5 | 0.9933 | 6 | 0.9978 | 6 | 0.9972 | 6 | 0.9976 | 6 |
| serine | 0.995 | 6 | 0.9961 | 5 | 0.9905 | 6 | 0.9702 | 6 | 0.9899 | 6 |
| sn-glycerol 3-phosphate | 0.9971 | 5 | 0.9937 | 6 | 0.9974 | 6 | 0.9937 | 6 | 0.9663 | 5 |
| succinic acid | 0.9806 | 5 | 0.9969 | 6 | 0.9941 | 6 | 0.9308 | 4 [§] | 0.7423* | 6 |
| taurine | 0.9921 | 6 | 0.9987 | 5 | 0.9989 | 5 | 0.9938 | 5 | 0.9996 | 4 [§] |
| taurocholic acid | 0.9953 | 9 | 0.9502 | 9 | 0.9994 | 10 | 0.9961 | 9 | 0.997 | 10 |
| thiamin | 0.8603* | 6 | 0.9922 | 6 | 0.9414 | 6 | 0.9372 | 8 | 0.8880* | 5 |
| thiamin monophosphate | 0.9784 | 6 | 0.9797 | 5 | 0.9643 | 6 | 0.9955 | 6 | 0.9954 | 6 |
| threonine | 0.9974 | 6 | 0.9947 | 5 | 0.9893 | 6 | 0.9978 | 6 | 0.9802 | 6 |
| thymidine | 0.9885 | 6 | 0.9964 | 6 | 0.9977 | 7 | 0.9973 | 7 | 0.9932 | 7 |
| tryptamine | 0.9981 | 9 | 0.9995 | 9 | 0.999 | 8 | 0.9981 | 8 | 0.9997 | 8 |
| tryptophan | 0.9998 | 9 | 0.9992 | 8 | 0.9999 | 9 | 0.9998 | 9 | 0.996 | 8 |
| tyrosine | 0.9696 | 6 | 0.9964 | 5 | 0.9878 | 6 | 0.998 | 6 | 0.977 | 7 |
| uracil | 0.9835 | 6 | --- | 2 [§] | 0.9822 | 6 | 0.9801 | 6 | --- | 2 [§] |
| uridine 5'-monophosphate | 0.9947 | 6 | 0.994 | 6 | 0.9417 | 5 | 0.9667 | 6 | 0.9994 | 5 |
| xanthine | 0.9866 | 6 | 0.9978 | 6 | 0.9992 | 6 | 0.9877 | 6 | 0.9473 | 6 |
| xanthosine | 1 | 6 | 0.9953 | 6 | 0.9998 | 6 | 0.9995 | 9 | 0.9679 | 8 |

* r^2 of the calibration curve was too low to use.

§ too few high-quality points to make a calibration curve.

¹Nicotinamide adenine dinucleotide phosphate. ²Nicotinamide adenine dinucleotide. ³ γ -aminobutyric acid.

⁴Dimethylsulfoniopropionate.

Table S6. Comparison of response factors in the TIS, VSW, Rpom, and Mp matrices to the response factor in MQ. The ratio of the response factor in each of the four matrices versus the response factor in Milli-Q water is shown as well as the *p*-value for the significance of the difference between the response factors based on the wild bootstrap method.¹ TIS/MQ, VSW/MQ, Rpom/MQ, and Mp/MQ refer to the response factor ratios of each matrix over the response factor in the MQ matrix. (--- indicates that the calibration curve was of insufficient quality to compare the response factors in that matrix or, in cases where just the *p*-value is missing, that no *p*-value could be calculated due to insufficient calibration points in common between the compared calibration curves. The ratio and the *p*-value of analytes that have a significant *p*-value (*p* < 0.05) in a treatment are in bold) Also, note that in some cases the ratio is fairly far from 1 but the *p*-value indicates a non-significant difference. This is due to the fact that this statistical method accounts for heteroscedasticity which often results in greater variability at higher concentrations which are also the concentrations that strongly influence the slope of the calibration curve.

(1) Estevez-Perez, G.; Andrade, J. M.; Wilcox, R. R. *Anal. Chem.* **2016**, *88*, 2289–2295.

| Compound | TIS/MQ | <i>P</i> -value | VSW/MQ | <i>P</i> -value | Rpom/MQ | <i>P</i> -value | Mp/MQ | <i>P</i> -value |
|----------------------------------|------------|-----------------|------------|-----------------|------------|-----------------|------------|-----------------|
| (6R)-5,6,7,8-tetrahydrobiopterin | --- | --- | --- | --- | --- | --- | --- | --- |
| 1-deoxy-D-xylulose-5-phosphate | 0.8 | 0.032 | 0.8 | 0.032 | 0.3 | 0.118 | 0.2 | 0.01 |
| 2,3-dihydroxybenzoic acid | 1 | 0.816 | 1.8 | 0.028 | 1.6 | 0.192 | 1.7 | 0.056 |
| 2,3-dihydroxypropane-1-sulfonate | 0.8 | 0.294 | 0.7 | 0.216 | 0.3 | 0.002 | 0.2 | 0.008 |
| 3-mercaptopropionic acid | 1.1 | 0.564 | 1.4 | 0.018 | 1.4 | 0.042 | 0.6 | 0.004 |
| 4-aminobenzoic acid | 0.9 | 0.006 | 0.9 | 0.006 | 0.9 | 0.004 | 0.9 | 0.004 |
| 4-hydroxybenzoic acid | 1.1 | 0.146 | 1.4 | 0.038 | 1.5 | 0.03 | 1.3 | 0.194 |
| 5-methylthioadenosine | --- | --- | --- | --- | --- | --- | --- | --- |
| 6-phosphogluconic acid | 0.7 | 0.11 | 0.8 | 0.082 | 0.4 | 0.098 | 0.4 | 0.166 |
| acetyltaurine | 0.9 | 0.504 | 1 | 0.756 | 0.4 | 0 | 0.4 | 0 |
| adenine | 1 | 0.94 | 1.1 | 0.752 | 1.2 | 0.524 | 1.1 | 0.628 |
| adenosine | 0.9 | 0.57 | 0.8 | 0.206 | --- | --- | 0.8 | 0.256 |
| adenosine 5'-monophosphate | 0.9 | 0.552 | 1.2 | 0.358 | 0.5 | 0.354 | 0.3 | 0.154 |
| alpha-ketoglutaric acid | 0.9 | 0.736 | 1.1 | 0.86 | 0.7 | 0.572 | 0.8 | 0.644 |
| arginine | 0.8 | 0.256 | 0.7 | 0.088 | 0.5 | 0.104 | --- | --- |
| aspartic acid | 1.1 | 0.61 | 1.1 | 0.654 | 0.7 | 0.184 | --- | --- |
| betaine | 1 | 0.964 | 1.1 | 0.6 | 1 | 0.712 | 1.1 | 0.544 |
| biotin | 1 | 0.972 | 1 | 0.806 | 1 | 0.246 | 1.1 | 0.776 |
| caffeine | 1 | 0.982 | 1 | 0.98 | 1 | 0.954 | 1 | 0.996 |
| chitobiose | 0.8 | 0.028 | 0.9 | 0.054 | 0.9 | 0.092 | 0.8 | 0.076 |
| chitotriose | 0.9 | 0.376 | 1 | 0.648 | 0.8 | 0.458 | 0.8 | 0.676 |
| choline | 0.9 | 0.592 | 1 | 0.736 | 1 | 0.76 | 1 | 0.73 |
| ciliatine | 1 | 0.638 | 1 | 0.674 | 0.8 | 0.334 | 0.8 | 0.31 |
| citric acid | 1.2 | 0.588 | 1.3 | 0.578 | 1 | 0.696 | 1.9 | 0.172 |
| citrulline | 1.1 | 0.528 | 1 | 0.632 | 0.8 | 0.192 | 0.9 | 0.552 |
| cyanocobalamin | 0.7 | 0.006 | 0.9 | 0.038 | 0.8 | 0.014 | 0.8 | 0.006 |
| cysteine | 1.2 | 0.166 | 0.9 | 0.2 | 1 | 0.796 | 1.1 | 0.23 |
| cytosine | 1.2 | 0.042 | 1 | 0.752 | 0.8 | 0.046 | 0.7 | 0.004 |
| D-(-)3-phosphoglyceric acid | 1 | 0.45 | 0.9 | 0.182 | 0.4 | 0.01 | 0.5 | 0.016 |
| D-glucosamine 6-phosphate | 0.6 | 0.306 | 0.8 | 0.51 | 0.6 | 0.478 | 0.2 | --- |

| | | | | | | | | |
|-----------------------------------|------------|--------------|------------|--------------|------------|--------------|------------|--------------|
| D-ribose 5-phosphate | 0.8 | 0.6 | 1 | 0.998 | 0.9 | 0.826 | 0.7 | 0.428 |
| desthiobiotin | --- | --- | --- | --- | --- | --- | --- | --- |
| dihydroxyacetone phosphate | 1.2 | 0.154 | 1.2 | 0.118 | 1.1 | 0.532 | 1.2 | 0.144 |
| DMSF ⁴ | 1 | 0.812 | 0.7 | 0.192 | 0.3 | 0.206 | 0.4 | 0.198 |
| ectoine | 1 | 0.84 | 1.1 | 0.572 | 1.1 | 0.564 | 1.1 | 0.456 |
| folic acid | 1 | 0.612 | 1.1 | 0.102 | 1.1 | 0.402 | 1.3 | 0.002 |
| fosfomycin | 0.8 | 0.296 | 0.8 | 0.186 | 0.3 | 0.096 | 0.2 | 0.004 |
| fumaric acid | 1.3 | 0.478 | --- | --- | 1.1 | 0.732 | 0.9 | 0.63 |
| GABA ³ | 1.1 | 0.564 | 1.1 | 0.648 | 1.1 | 0.564 | 1.2 | 0.43 |
| glucose 6-phosphate | 1 | 0.896 | 0.8 | 0.444 | 0.5 | 0.412 | 0.4 | 0.324 |
| glutamic acid | 1 | 0.468 | 1 | 0.604 | 0.6 | 0.016 | 0.8 | 0.02 |
| glutamine | 1.2 | 0.408 | 1 | 0.842 | 0.8 | 0.516 | 1 | 0.888 |
| glutathione | 1 | 0.728 | 1 | 0.838 | 0.3 | 0.142 | 0.4 | 0.106 |
| glutathione oxidized | 0.9 | 0.074 | 1.3 | 0.016 | 0.5 | 0.042 | 0.4 | 0.02 |
| glyphosate | 0.8 | 0.53 | 0.6 | 0.098 | 1.1 | 0.098 | 1.2 | 0.096 |
| guanine | 0.9 | 0.604 | 1 | 0.816 | 0.8 | 0.532 | --- | --- |
| guanosine | 1 | 0.628 | 1.1 | 0.106 | 1 | 0.398 | 1 | 0.534 |
| indole 3-acetic acid | 0.4 | 0.014 | 1 | 0.19 | 1 | 0.146 | 0.6 | 0.006 |
| inosine | 0.8 | 0.176 | 0.9 | 0.564 | 0.6 | 0.068 | 0.6 | 0.066 |
| inosine 5'-monophosphate | 0.9 | 0.57 | 1 | 0.734 | 1 | 0.664 | 1 | 0.986 |
| (iso)leucine | 1 | 0.814 | 0.9 | 0.608 | 0.8 | 0.536 | --- | --- |
| kynurenine | 0.9 | 0.03 | 1.1 | 0.042 | 0.9 | 0.07 | 1 | 0.192 |
| malic acid | 1 | 0.834 | 1 | 0.57 | 0.3 | 0 | 0.3 | 0.002 |
| methionine | 1.1 | 0.464 | 1 | 0.808 | 0.3 | 0.156 | 0.3 | 0.138 |
| muramic acid | 0.9 | 0.346 | 1.3 | 0.274 | 1.3 | 0.178 | 1.2 | 0.4 |
| N-acetylglucosamine | 1 | 0.988 | 1 | 0.884 | 0.9 | 0.562 | 0.9 | 0.782 |
| N-acetylglutamic acid | 0.7 | 0.026 | 1 | 0.582 | 0.7 | 0.094 | 0.7 | 0.028 |
| N-acetylmuramic acid | 0.9 | 0.53 | 1.1 | 0.566 | 1 | 0.87 | 1 | 0.682 |
| NAD ² | 0.8 | 0.528 | 1 | 0.978 | 0.5 | 0.4 | 0.4 | 0.288 |
| NADP ¹ | 0.6 | 0.188 | 0.8 | 0.52 | 0.4 | 0.09 | 0.3 | 0.124 |
| ornithine | 0.9 | 0.064 | 0.7 | 0.014 | 0.4 | 0.012 | 0.3 | 0.022 |
| orotic acid | 1 | 0.824 | 1 | 0.586 | 0.6 | 0.094 | 0.5 | 0.03 |
| pantothenic acid | 0.9 | 0.032 | 1.2 | 0.03 | 1 | 0.674 | 1 | 0.478 |
| phenylalanine | 0.9 | 0.076 | 1 | 0.254 | 1 | 0.582 | --- | --- |
| phosphoenolpyruvate | 0.9 | 0.204 | 0.8 | 0.17 | --- | --- | --- | --- |
| proline | 1.1 | 0.566 | 1 | 0.932 | 1.1 | 0.516 | 1.2 | 0.466 |
| putrescine | 1.2 | 0.044 | 1.2 | 0.064 | 1.4 | 0.016 | 1.3 | 0.018 |
| pyridoxine | 1.1 | 0.62 | 1.1 | 0.568 | 1 | 0.842 | 0.9 | 0.61 |
| riboflavin | 1 | 0.392 | 1 | 0.632 | 0.9 | 0.274 | 1 | 0.734 |
| S-(1,2-dicarboxyethyl)glutathione | 0.9 | 0.598 | 0.8 | 0.524 | 0.7 | 0.344 | 0.7 | 0.444 |
| S-(5'-adenosyl)-L-homocysteine | 1.1 | 0.098 | 1.1 | 0.052 | --- | --- | --- | --- |
| S-adenosyl methionine | 0.8 | 0.338 | 0.8 | 0.284 | 0.8 | 0.202 | 1 | 0.938 |
| sarcosine | 1.3 | 0.108 | 1.1 | 0.124 | 1.1 | 0.132 | 1.2 | 0.092 |
| serine | 1.2 | 0.15 | 1 | 0.524 | 0.8 | 0.208 | 0.9 | 0.576 |
| sn-glycerol 3-phosphate | 0.8 | 0.102 | 0.7 | 0.084 | 0.4 | 0.05 | 0.4 | 0.128 |
| succinic acid | 1.1 | 0.876 | 1.1 | 0.83 | 0.8 | 0.458 | 0.8 | 0.552 |
| taurine | 1.1 | 0.576 | 1.1 | 0.57 | 0.7 | 0.198 | 1 | 0.89 |
| taurocholic acid | 0.8 | 0.308 | 1.1 | 0.122 | 1 | 0.576 | 0.5 | 0 |
| thiamin | 1.1 | 0.162 | 1.1 | 0.222 | 1.2 | 0.094 | 1.4 | 0.03 |
| thiamin monophosphate | 1.3 | 0.45 | 1.1 | 0.594 | 1 | 0.682 | 1.1 | 0.542 |
| threonine | 1.1 | 0.566 | 1.1 | 0.584 | 0.9 | 0.592 | 1 | 0.96 |
| thymidine | 1 | 0.588 | 1.3 | 0.054 | 1.1 | 0.27 | 1 | 0.976 |
| tryptamine | 0.8 | 0.006 | 1 | 0.536 | 1 | 0.648 | 0.9 | 0.122 |

| | | | | | | | | |
|--------------------------|-----|-------|-----|-------|------------|--------------|------------|--------------|
| tryptophan | 0.9 | 0.036 | 1 | 0.874 | 0.9 | 0.05 | 0.6 | 0.01 |
| tyrosine | 1 | 0.886 | 1 | 0.994 | 0.9 | 0.274 | 0.8 | 0.09 |
| uracil | --- | --- | 1 | 0.86 | 1.1 | 0.612 | --- | --- |
| uridine 5'-monophosphate | 0.9 | 0.08 | 0.8 | 0.07 | 0.4 | 0.056 | 0.3 | 0.032 |
| xanthine | 1 | 0.23 | 1 | 0.916 | 0.9 | 0.022 | 0.8 | 0.018 |
| xanthosine | 1 | 0.282 | 1.1 | 0.396 | 0.9 | 0.132 | 0.5 | 0.036 |

¹Nicotinamide adenine dinucleotide phosphate. ²Nicotinamide adenine dinucleotide. ³ γ -aminobutyric acid.

⁴Dimethylsulfoniopropionate.

Table S7. Average percentage of metabolite retained after solid phase extraction and one standard deviation of triplicate extractions. Blanks (---) are metabolites that could not be quantified because the calibration curve was not of sufficiently high quality. ‘n.r.’ indicates metabolites that were not retained in a given treatment. The final column indicates whether extraction efficiency in the culture treatments compared to the other treatments is about the same (‘equal’) or higher (‘culture’) or lower (‘non-culture’). Compounds are presented according to their retention time (RT). An asterisk (*) denotes metabolites who were spiked into the matrices at 3x the concentration of the other compounds.

| Compound | RT (min) | MQ | TIS | VSW | Rpom | Mp | Culture vs Non-Culture |
|-------------------------------------------|----------|---------|---------|---------|---------|---------|------------------------|
| | | Percent | Percent | Percent | Percent | Percent | |
| taurocholic acid | 17 | 64±17 | 45±22 | 93±5 | 87±22 | 115±10 | equal |
| indole 3-acetic acid | 11.5 | 65±3 | 27±27 | 17±6 | 48±3 | 51±21 | equal |
| desthiobiotin | 9.9 | 7.8±5.0 | 2.1±1.2 | 6.5±3.2 | 4.6±4.0 | 15±6 | equal |
| 3-mercaptopropionic acid | 9.7 | 93±6 | 85±36 | 89±8 | 61±27 | 95±25 | equal |
| cyanocobalamin | 9.5 | 83±11 | 75±13 | 79±4 | 96±4 | 69±41 | equal |
| riboflavin | 9.1 | 106±4 | 96±2 | 88±0.2 | 125±5 | 71±42 | equal |
| biotin | 8.8 | 57±5 | 53±11 | 53±4 | 77±3 | 45±30 | equal |
| caffeine | 8.4 | 23±2 | 26±4 | 24±4 | 20±5 | 18±7 | equal |
| 2,3-dihydroxybenzoic acid | 8.3 | 95±7 | 92±7 | 101±8 | 114±5 | 52±30 | equal |
| folic acid | 8.1 | 80±19 | 68±22 | 41±13 | 42±8 | 39±18 | equal |
| 4-hydroxybenzoic acid | 7.6 | 92±5 | 83±7 | 88±8 | 109±5 | 121±11 | culture |
| 5-methylthioadenosine | 6.2 | 84±15 | 86±6 | 81±2 | 78±4 | 84±4 | equal |
| tryptophan | 5.7 | 74±6 | 57±33 | 47±9 | 71±2 | 129±41 | equal |
| tryptamine | 5.7 | 72±5 | 41±34 | 21±6 | 43±1 | 64±9 | equal |
| 4-aminobenzoic acid | 5.2 | 45±10 | 40±7 | 19±2 | 10±4 | 1.8±2.2 | non-culture |
| pantothenic acid | 3.9 | 50±9 | 60±2 | 52±4 | 95±3 | 100±8 | culture |
| thymidine | 3.8 | 51±14 | 49±5 | 53±4 | 85±4 | 82±12 | culture |
| kynurenine | 3.4 | 57±9 | 67±6 | 42±4 | 32±7 | 0.7±0.5 | non-culture |
| xanthosine | 3.3 | 11±2 | 12±0.6 | 10±2 | 46±4 | 97±38 | culture |
| phenylalanine | 3.2 | 22±1 | 47±3 | 40±2 | 82±3 | --- | culture |
| guanosine | 2.7 | 6±0.9 | 7.6±0.2 | 7.7±0.5 | 32±3 | 38±1 | culture |
| inosine* | 2.7 | 8.7±2.4 | 8.7±0.8 | 8.1±1.4 | 41±1 | 39±2 | culture |
| inosine 5'-monophosphate | 2.5 | n.r. | n.r. | n.r. | n.r. | n.r. | --- |
| <i>N</i> -acetylglutamic acid* | 2.5 | 1.0±0.1 | 2.0±0.8 | 1.1±1.0 | 6.6±0.6 | 4.8±1.1 | culture |
| xanthine* | 2.5 | 0.3±0.2 | 0.2±0.2 | 0.3±0.2 | 1.4±0.2 | 2.8±4.9 | equal |
| fumaric acid* | 2.5 | 0.5±0.9 | n.r. | --- | n.r. | n.r. | --- |
| <i>N</i> -acetylmuramic acid | 2.4 | 3.2±2.7 | 0.9±1.5 | 2.8±2.7 | 16±4 | 18±2 | culture |
| (6R)-5,6,7,8-tetrahydrobiopterin* | 2.4 | 0.8±0.1 | n.r. | n.r. | n.r. | n.r. | --- |
| succinic acid* | 2.4 | n.r. | n.r. | n.r. | --- | --- | --- |
| adenosine* | 2.4 | 3.3±0.3 | 5.2±0.5 | 6.5±1.9 | 14±0.6 | 11±0.5 | culture |
| <i>S</i> -(1,2-dicarboxyethyl)glutathione | 2.3 | n.r. | n.r. | n.r. | n.r. | --- | --- |

| | | | | | | | |
|-----------------------------------|-----|---------|---------|---------|----------|---------|---------|
| (iso)leucine | 2.3 | 0.5±0.8 | 3.9±3.4 | 2.9±2.6 | 7.8±2.8 | 18±32 | equal |
| adenosine 5' monophosphate* | 2.3 | 0.3±0.0 | 0.3±0.2 | 0.2±0.0 | 1.7±0.7 | 1.0±1.5 | equal |
| Tyrosine* | 2.3 | 1.2±0.3 | 2.9±0.3 | 2.1±0.3 | 6.0±1.2 | 8.1±3.8 | culture |
| alpha-ketoglutaric acid* | 2.3 | n.r. | n.r. | n.r. | 0.8±1.4 | --- | --- |
| orotic acid* | 2.3 | 0.1±0.1 | n.r. | n.r. | n.r. | n.r. | --- |
| S-(5'-adenosyl)-L-homocysteine* | 2.3 | 12±3 | 41±5 | 43±6 | 80±10 | 54±9 | culture |
| uracil* | 2.3 | n.r. | --- | n.r. | 1.2±2.1 | --- | --- |
| citric acid* | 2.2 | n.r. | n.r. | 0.9±1.5 | n.r. | 0.6±0.5 | --- |
| NADP ¹ | 2.2 | 0.5±0.9 | 1.2±2.1 | 5.5±2.2 | n.r. | 4.6±8.0 | --- |
| NAD ² | 2.2 | 7.8±4.6 | 14±3 | 21±9 | 74±17 | 25±29 | culture |
| uridine 5'-monophosphate* | 2.2 | 0.2±0.3 | n.r. | n.r. | n.r. | n.r. | --- |
| Acetyltaurine* | 2 | n.r. | n.r. | n.r. | n.r. | n.r. | --- |
| dihydroxyacetone phosphate | 2 | n.r. | n.r. | n.r. | n.r. | --- | --- |
| fosfomycin* | 2 | --- | --- | n.r. | n.r. | n.r. | --- |
| phosphoenolpyruvate | 2 | n.r. | n.r. | n.r. | n.r. | --- | --- |
| methionine | 2 | 0.3±0.1 | 0.3±0.5 | n.r. | n.r. | --- | --- |
| 2,3-dihydroxypropane-1-sulfonate* | 1.9 | n.r. | 0.1±0.2 | 0.6±1.0 | --- | --- | --- |
| 1-deoxy-D-xylulose-5-phosphate* | 1.9 | 0.1±0.1 | n.r. | n.r. | n.r. | n.r. | --- |
| betaine | 1.9 | n.r. | n.r. | n.r. | n.r. | 0.3±0.2 | --- |
| glyphosate | 1.9 | n.r. | n.r. | 15±4 | 56±49 | --- | equal |
| glutathione | 1.9 | n.r. | n.r. | 1.1±2.0 | 1.2±2.1 | 1.3±2.3 | --- |
| malic acid | 1.9 | n.r. | n.r. | 0.7±0.6 | n.r. | n.r. | --- |
| guanine | 1.9 | n.r. | n.r. | n.r. | 18.0±6.9 | 0.8±1.0 | --- |
| glutathione oxidized | 1.9 | n.r. | 3.5±0.8 | 1.5±2.6 | 10±2 | --- | --- |
| pyridoxine* | 1.9 | 2.7±0.0 | 7.5±0.7 | 6.8±0.9 | 32±12 | 31±1 | culture |
| adenine* | 1.8 | n.r. | n.r. | n.r. | 0.6±1.1 | --- | --- |
| D-ribose 5-phosphate | 1.8 | n.r. | n.r. | 0.1±0.1 | 2.8±4.8 | 0.7±1.3 | --- |
| sn-glycerol 3-phosphate | 1.8 | n.r. | n.r. | n.r. | n.r. | n.r. | --- |
| D-(-)3-phosphoglyceric acid | 1.8 | n.r. | n.r. | 1.7±3.0 | n.r. | --- | --- |
| 6-phosphogluconic acid | 1.8 | 0.6±1.1 | n.r. | n.r. | 2.0±3.5 | 6.9±6.0 | --- |
| ectoine | 1.8 | n.r. | n.r. | n.r. | n.r. | n.r. | --- |
| glucose 6-phosphate | 1.8 | n.r. | n.r. | n.r. | n.r. | n.r. | --- |
| proline | 1.8 | n.r. | n.r. | n.r. | n.r. | 0.2±0.3 | --- |
| chitobiose | 1.8 | n.r. | n.r. | 0.3±0.5 | 5.6±9.8 | --- | --- |
| chitotriose | 1.8 | n.r. | 5±3 | 6±2 | 38±9 | 15±14 | culture |
| choline | 1.8 | n.r. | n.r. | n.r. | n.r. | n.r. | --- |
| GABA ³ | 1.8 | n.r. | n.r. | n.r. | n.r. | n.r. | --- |
| N-acetylglucosamine | 1.7 | n.r. | n.r. | n.r. | 1.3±2.2 | n.r. | --- |
| DMSP ⁴ * | 1.7 | n.r. | n.r. | n.r. | n.r. | 0.1±0.2 | --- |
| cytosine | 1.7 | n.r. | n.r. | n.r. | n.r. | n.r. | --- |
| cysteine | 1.7 | n.r. | n.r. | n.r. | 1.2±1.0 | 1.1±0.9 | --- |
| thiamin monophosphate | 1.7 | n.r. | n.r. | n.r. | n.r. | n.r. | --- |
| muramic acid | 1.7 | n.r. | n.r. | n.r. | n.r. | n.r. | --- |

| | | | | | | | |
|---------------------------|-----|---------|---------|---------|---------|---------|-----|
| taurine* | 1.7 | n.r. | n.r. | n.r. | 0.1±0.1 | --- | --- |
| ciliatine | 1.6 | n.r. | n.r. | n.r. | n.r. | n.r. | --- |
| citrulline | 1.6 | n.r. | n.r. | n.r. | 0.1±0.1 | n.r. | --- |
| glutamine | 1.6 | n.r. | n.r. | n.r. | n.r. | 1.8±3.2 | --- |
| sarcosine | 1.6 | n.r. | n.r. | n.r. | n.r. | 0.6±1.1 | --- |
| serine | 1.6 | n.r. | 0.2±0.3 | n.r. | n.r. | 1.3±1.2 | --- |
| threonine | 1.6 | n.r. | n.r. | n.r. | 0.4±0.0 | 0.8±1.4 | --- |
| aspartic acid | 1.6 | n.r. | n.r. | n.r. | n.r. | n.r. | --- |
| glutamic acid | 1.6 | n.r. | n.r. | n.r. | n.r. | n.r. | --- |
| thiamin* | 1.6 | --- | 2.0±1.0 | 2.0±0.5 | 1.8±3.1 | --- | --- |
| arginine | 1.6 | 0.1±0.1 | n.r. | n.r. | n.r. | --- | --- |
| D-glucosamine 6-phosphate | 1.5 | n.r. | n.r. | n.r. | n.r. | n.r. | --- |
| ornithine | 1.4 | n.r. | n.r. | n.r. | n.r. | n.r. | --- |
| S-adenosyl methionine | 1.4 | n.r. | n.r. | n.r. | 0.2±0.4 | n.r. | --- |
| putrescine | 1.3 | n.r. | n.r. | n.r. | 0.6±1.1 | n.r. | --- |

¹Nicotinamide adenine dinucleotide phosphate. ²Nicotinamide adenine dinucleotide. ³ γ -aminobutyric acid.

⁴Dimethylsulfoniopropionate.

Table S8. Relationship between molecular structure and PPL extraction efficiency. The retention time, extraction efficiency in Milli-Q water, and some structural information for each metabolite are provided in the table. Under structural information, we present the charge of each analyte at pH 2 (calculated, where possible, using SPARC (<http://archemcalc.com/>) an online computational tool for calculating molecular properties based on the structure of a molecule), the presence (+P) of a phosphate group, and the presence of an aromatic moiety. Blank (---) are compounds that were not aromatic and no charge could be calculated. 'n.r.' indicates metabolites that were not retained in a given treatment. Data are sorted by retention time (RT) of the metabolite.

| Compound | RT (min) | MQ Extraction Efficiency | Structural Information |
|-------------------------------------------|----------|--------------------------|------------------------|
| | | Percent | |
| taurocholic acid | 17 | 64±17 | Negative |
| indole 3-acetic acid | 11.5 | 65±3 | Aromatic/Neutral |
| desthiobiotin | 9.9 | 8±5 | Neutral |
| 3-mercaptopropionic acid | 9.7 | 93±6 | Neutral |
| cyanocobalamin | 9.5 | 83±11 | Aromatic |
| riboflavin | 9.1 | 106±4 | Aromatic/Positive |
| biotin | 8.8 | 57±5 | Neutral |
| caffeine | 8.4 | 23±2 | Aromatic/Neutral |
| 2,3-dihydroxybenzoic acid | 8.3 | 95±7 | Aromatic/Neutral |
| folic acid | 8.1 | 80±19 | Aromatic/Positive |
| 4-hydroxybenzoic acid | 7.6 | 92±5 | Aromatic/Neutral |
| 5-methylthioadenosine | 6.2 | 84±15 | Aromatic/Positive |
| tryptophan | 5.7 | 74±6 | Aromatic/Positive |
| tryptamine | 5.7 | 72±5 | Aromatic/Positive |
| 4-aminobenzoic acid | 5.2 | 45±10 | Aromatic/Positive |
| pantothenic acid | 3.9 | 50±9 | Neutral |
| thymidine | 3.8 | 51±14 | Aromatic/Neutral |
| kynurenine | 3.4 | 57±9 | Aromatic/Positive |
| xanthosine | 3.3 | 11±2 | Aromatic/Neutral |
| phenylalanine | 3.2 | 22±1 | Aromatic/Positive |
| guanosine | 2.7 | 6.2±0.9 | Aromatic/Positive |
| inosine | 2.7 | 8.7±2.4 | Aromatic/Positive |
| inosine 5'-monophosphate | 2.5 | n.r. | Aromatic + P |
| <i>N</i> -acetylglutamic acid | 2.5 | 1.0±0.1 | Neutral |
| xanthine | 2.5 | 0.3±0.2 | Aromatic/Neutral |
| fumaric acid | 2.5 | 0.5±0.9 | neutral |
| <i>N</i> -acetylmuramic acid | 2.4 | 3.2±2.7 | Neutral |
| (6R)-5,6,7,8-tetrahydrobiopterin | 2.4 | 0.8±0.1 | Positive |
| succinic acid | 2.4 | n.r. | neutral |
| adenosine | 2.4 | 3.3±0.3 | Aromatic/Positive |
| <i>S</i> -(1,2-dicarboxyethyl)glutathione | 2.3 | n.r. | --- |
| (iso)leucine | 2.3 | 0.5±0.8 | Positive |

| | | | |
|----------------------------------|-----|---------|---------------------|
| adenosine 5'-monophosphate | 2.3 | 0.3±0.0 | Aromatic + P |
| tyrosine | 2.3 | 1.2±0.3 | Aromatic/Positive |
| alpha-ketoglutaric acid | 2.3 | n.r. | Neutral |
| orotic acid | 2.3 | 0.1±0.1 | Neutral |
| S-(5'-adenosyl)-L-homocysteine | 2.3 | 12±3 | Aromatic/Positive |
| uracil | 2.3 | n.r. | Aromatic/Neutral |
| citric acid | 2.2 | n.r. | Neutral |
| NADP ¹ | 2.2 | 0.5±0.9 | Aromatic + P |
| NAD ² | 2.2 | 7.8±4.6 | Aromatic |
| uridine 5'-monophosphate | 2.2 | 0.2±0.3 | Aromatic + P |
| acetyltaurine | 2 | n.r. | Negative |
| dihydroxyacetone phosphate | 2 | n.r. | + P |
| fosfomycin | 2 | --- | Phosphonate |
| phosphoenolpyruvate | 2 | n.r. | --- |
| methionine | 2 | 0.3±0.1 | Positive/Zwitterion |
| 2,3-dihydroxypropane-1-sulfonate | 1.9 | n.r. | + S |
| 1-deoxy-D-xylulose-5-phosphate | 1.9 | 0.1±0.1 | + P |
| betaine | 1.9 | n.r. | --- |
| glyphosate | 1.9 | n.r. | Negative/Zwitterion |
| glutathione | 1.9 | n.r. | Positive/Zwitterion |
| malic acid | 1.9 | n.r. | Neutral |
| guanine | 1.9 | n.r. | Aromatic/Positive |
| glutathione oxidized | 1.9 | n.r. | --- |
| pyridoxine | 1.9 | 2.7±0.0 | Aromatic/Positive |
| adenine | 1.8 | n.r. | Aromatic/Positive |
| D-ribose 5-phosphate | 1.8 | n.r. | Sugar + P |
| sn-glycerol 3-phosphate | 1.8 | n.r. | + P |
| D-(-)3-phosphoglyceric acid | 1.8 | n.r. | + P |
| 6-phosphogluconic acid | 1.8 | 0.6±1.1 | Sugar + P |
| ectoine | 1.8 | n.r. | Positive |
| glucose 6-phosphate | 1.8 | n.r. | Sugar + P |
| proline | 1.8 | n.r. | Positive |
| chitobiose | 1.8 | n.r. | Sugar/Neutral |
| chitotriose | 1.8 | n.r. | Sugar/charge? |
| choline | 1.8 | n.r. | Positive |
| GABA ³ | 1.8 | n.r. | Positive/Zwitterion |
| N-acetylglucosamine | 1.7 | n.r. | Sugar/Neutral |
| DMSF ⁴ | 1.7 | n.r. | --- |
| cytosine | 1.7 | n.r. | Aromatic/Positive |
| cysteine | 1.7 | n.r. | Positive/Zwitterion |
| thiamin monophosphate | 1.7 | n.r. | Aromatic + P |
| muramic acid | 1.7 | n.r. | Positive |
| taurine | 1.7 | n.r. | Positive/Zwitterion |

| | | | |
|---------------------------|-----|---------|---------------------|
| ciliatine | 1.7 | n.r. | Positive/Zwitterion |
| citrulline | 1.7 | n.r. | Positive/Zwitterion |
| glutamine | 1.6 | n.r. | Positive/Zwitterion |
| sarcosine | 1.6 | n.r. | Positive/Zwitterion |
| serine | 1.6 | n.r. | Positive/Zwitterion |
| threonine | 1.6 | n.r. | Positive/Zwitterion |
| aspartic acid | 1.6 | n.r. | Positive/Zwitterion |
| glutamic acid | 1.6 | n.r. | Positive/Zwitterion |
| thiamin | 1.6 | --- | Aromatic – 2+ |
| arginine | 1.6 | 0.1±0.1 | Positive |
| D-glucosamine 6-phosphate | 1.5 | n.r. | Positive |
| ornithine | 1.4 | n.r. | Positive |
| S-adenosyl methionine | 1.4 | n.r. | --- |
| putrescine | 1.3 | n.r. | Positive |

¹Nicotinamide adenine dinucleotide phosphate. ²Nicotinamide adenine dinucleotide. ³γ-aminobutyric acid.

⁴Dimethylsulfoniopropionate.

Table S9. Molecular parameters considered when examining metabolites that showed improved extraction efficiencies in culture matrices. The metabolites in this table had extraction efficiencies of at least 1% and are categorized based on whether their extraction efficiency was higher, lower, or the same between the culture and non-culture treatments. An asterisk (*) indicates the extraction efficiency was higher in culture matrices, a caret (^) indicates a lower extraction efficiency in the culture matrices, and no mark indicates there was no statistical difference in the extraction efficiency between the culture and non-culture matrices. Columns 2-6 indicate the fractional abundance of charge states of each metabolite at pH 2. Polar Surface Area refers to the surface area of all the polar atoms (i.e. oxygen, nitrogen) in the molecule, and log K_{oc} is the partitioning coefficient at pH 2 between water and generic organic carbon. The charge composition and log K_{oc} at pH 2 were calculated using SPARC (<http://archemcalc.com/>).

| Metabolite | 2+ | + | neutral | zwitterion | negative (-) | Molecular Weight | Polar Surface Area | log K _{oc} (pH 2) |
|---------------------------------|------|------|---------|------------|--------------|------------------|--------------------|----------------------------|
| pyridoxine* | 0 | 1 | 0 | 0 | 0 | 169.074 | 74 | -1.734 |
| S-(5'-adenosyl)-L-homocysteine* | 0.16 | 0.61 | 0 | 0.22 | 0 | 384.123 | 208 | -2.373 |
| N-acetylmuramic acid* | 0 | 0 | 0.96 | 0 | 0.04 | 293.111 | 146 | -3.922 |
| inosine* | 0 | 0.87 | 0 | 0.13 | 0 | 268.081 | 134 | -0.205 |
| guanosine* | 0 | 0.98 | 0 | 0.02 | 0 | 283.092 | 156 | -0.597 |
| phenylalanine* | 0 | 0.62 | 0 | 0.38 | 0 | 165.079 | 63 | -0.922 |
| xanthosine* | 0 | 0.21 | 0.78 | 0 | 0 | 284.076 | 154 | -1.876 |
| thymidine* | 0 | 0 | 1 | 0 | 0 | 242.09 | 103 | -2.727 |
| pantothenic acid* | 0 | 0 | 1 | 0 | 0 | 219.112 | 110 | -3.725 |
| 4-hydroxybenzoic acid* | 0 | 0 | 0.99 | 0 | 0 | 138.032 | 58 | 1.441 |
| tyrosine* | 0 | 0.67 | 0 | 0.33 | 0 | 181.074 | 84 | -1.099 |
| adenosine* | 0 | 0.97 | 0.02 | 0 | 0 | 267.097 | 140 | -1.851 |
| N-acetylglutamic acid* | 0 | 0 | 0.98 | 0 | 0.02 | 189.064 | 107 | -3.189 |
| NAD ²⁺ * | --- | --- | --- | --- | --- | 663.425 | 341 | --- |
| chitotriose* | --- | --- | --- | --- | --- | 501.483 | 294 | --- |
| 4-aminobenzoic acid^ | 0 | 0.7 | 0.27 | 0.02 | 0 | 137.048 | 63 | 0.806 |
| kynurenine^ | 0.15 | 0.55 | 0 | 0.3 | 0 | 208.085 | 106 | -1.26 |
| taurocholic acid | 0 | 0 | 0.02 | 0 | 0.98 | 515.292 | 156 | -1.553 |
| indole 3-acetic acid | 0 | 0 | 1 | 0 | 0 | 175.063 | 53 | 0.349 |
| desthiobiotin | 0 | 0 | 1 | 0 | 0 | 214.132 | 82 | -1.2 |
| 3-mercaptopropionate | 0 | 0 | 0.99 | 0 | 0 | 120.024 | 76 | -0.826 |
| cyanocobalamin | --- | --- | --- | --- | --- | 1355.365 | 488 | --- |
| riboflavin | 0 | 0.77 | 0.21 | 0 | 0 | 376.138 | 159 | --- |
| biotin | 0 | 0 | 1 | 0 | 0 | 244.088 | 107 | -2.367 |
| caffeine | 0 | 0 | 0.99 | 0 | 0 | 194.08 | 58 | 0.256 |
| 2,3-dihydroxybenzoic acid | 0 | 0 | 0.91 | 0 | 0.09 | 154.027 | 78 | 1.938 |
| folic acid | 0.01 | 0.71 | 0 | 0.27 | 0 | 441.14 | 210 | -0.076 |
| 5-methylthioadenosine | 0 | 0.97 | 0.02 | 0 | 0 | 297.09 | 145 | -1.342 |
| tryptophan | 0 | 0.75 | 0 | 0.25 | 0 | 204.09 | 79 | -0.94 |
| tryptamine | 0 | 1 | 0 | 0 | 0 | 160.1 | 42 | -0.548 |
| xanthine | 0 | 0.04 | 0.95 | 0 | 0 | 152.033 | 95 | -1.039 |
| (iso)leucine | 0 | 0.7 | 0 | 0.3 | 0 | 131.095 | 63 | -0.984 |
| adenosine 5'-monophosphate | 0 | 0.12 | 0 | 0.85 | 0.02 | 347.063 | 196 | -2.476 |
| glyphosate | 0 | 0 | 0 | 0.6 | 0.39 | 169.014 | 117 | -3.043 |

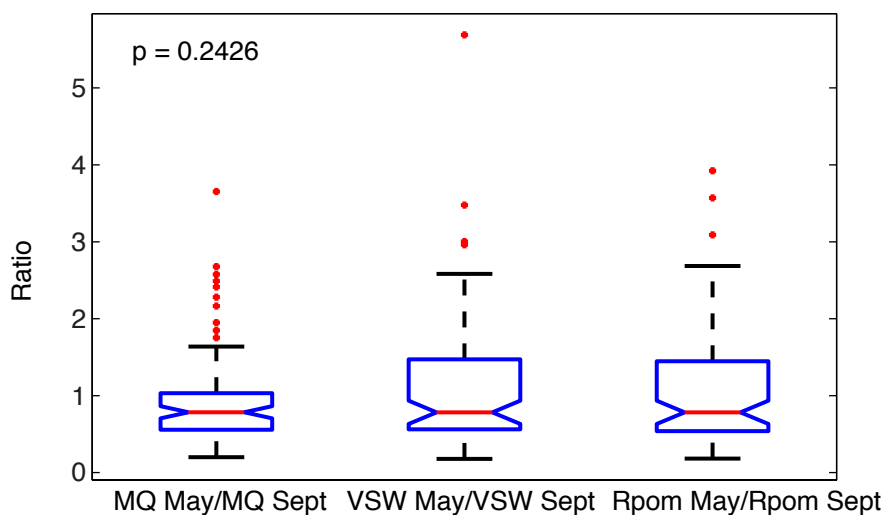


Figure S1. Comparison of instrument response factors (calibration curve slopes) made in varying matrix concentrations (May) and constant matrix concentrations (September). The variability of the ratio (May calibration slope/Sept calibration slope) is plotted for three matrices. Red dots ratios that are outliers. The variability of the ratios is not significantly different according to an ANOVA between the matrices and thus is attributed to instrument variability rather than the calibration curve matrix. See Text S1 for further details.

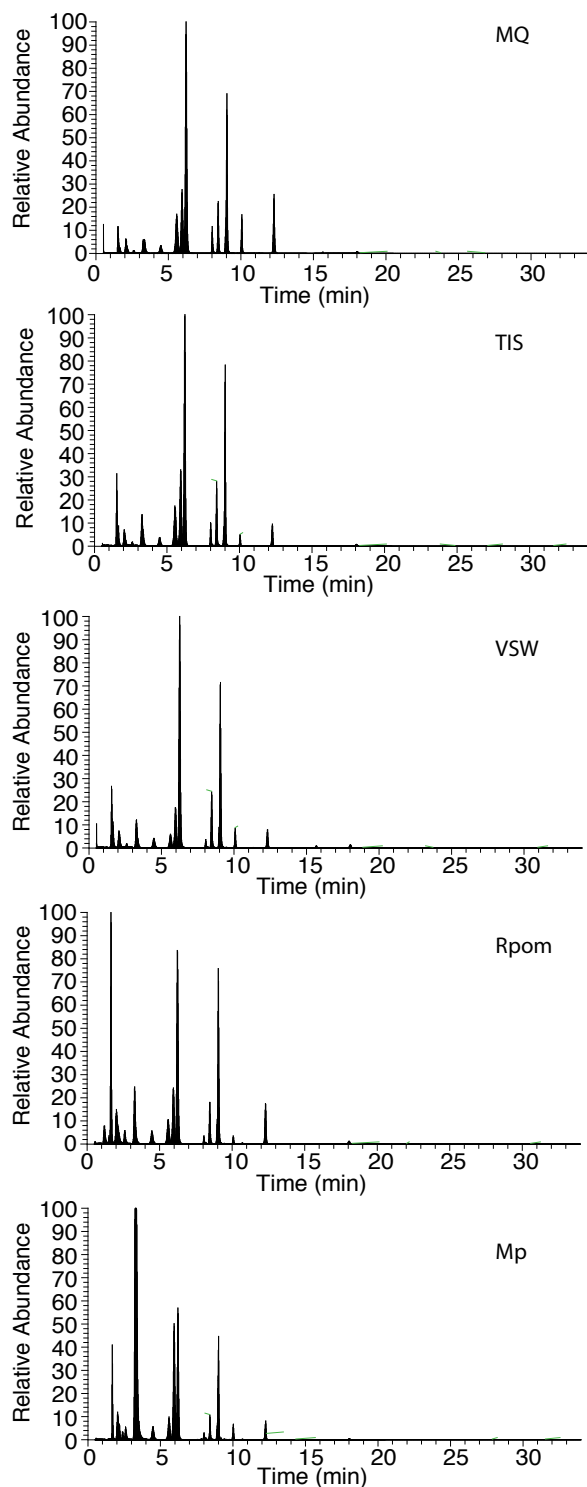


Figure S2. Composite chromatograms combine all of the scans over the course of a single sample run. An example is shown for each matrix type. Peaks represent only metabolites whose parent mass was scanned for over a designated retention time window and whose specified fragment was then detected. Contaminants do not appear in these chromatograms. Our method is optimized to reduce sample exposure to plasticizers so these contaminants should be minimal.

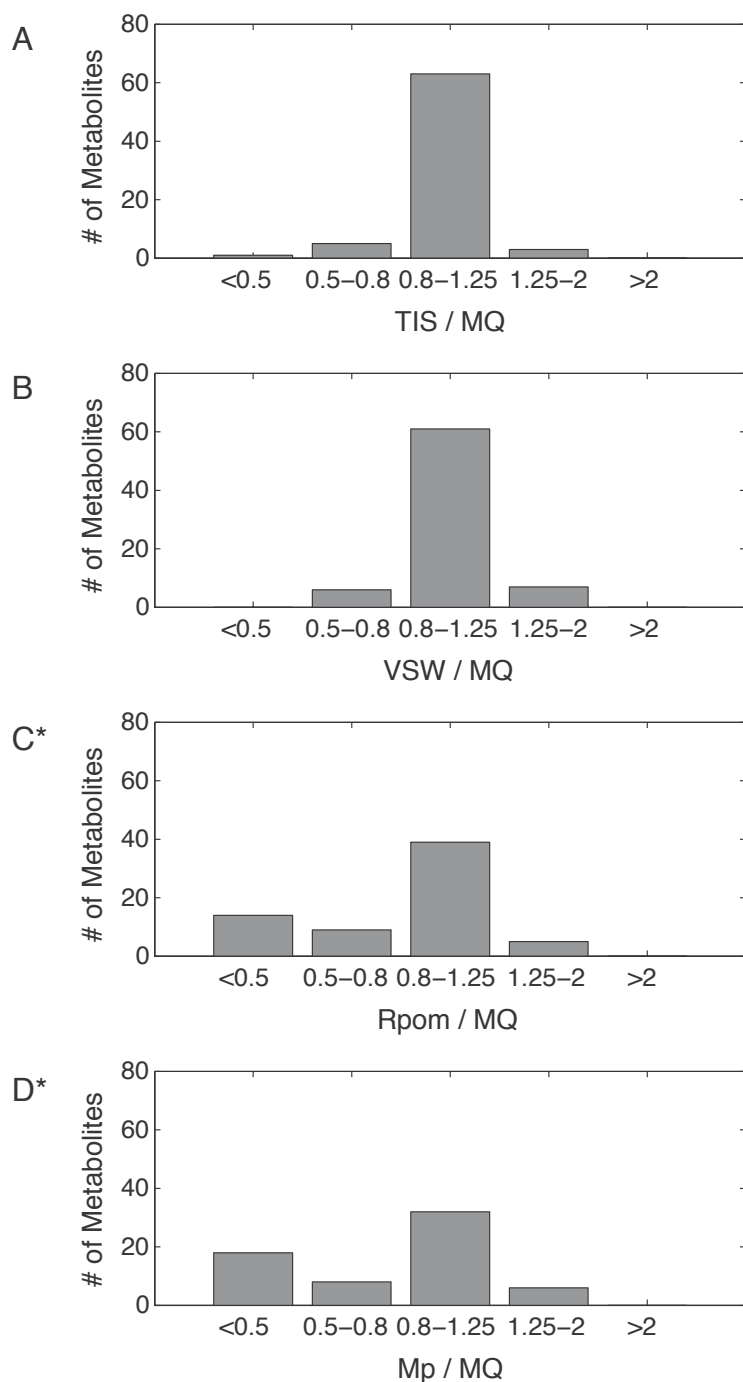


Figure S3. The ratios of the response factor of each analyte in the given matrix versus the response factor of the analyte in Milli-Q water are plotted to show the distribution of ratios in each matrix. To compare whether ionization suppression or enhancement was a significantly bigger factor in any of the matrices a Kolmogorov-Smirnov test was used to compare the distributions of ratios. This test found that the group without an asterisk (which is mostly grouped around a ratio of 1) and the group marked with an asterisk (which shows more reduced ratios) had significantly different distributions.

Chapter 3.

Evidence for quorum sensing and differential metabolite production by a marine bacterium in response to DMSP

This chapter was published in the ISME Journal by the Nature Publishing Group under a Creative Commons license.

Johnson, W. M., M. C. Kido Soule, and E. B. Kujawinski. 2016. Evidence for quorum sensing and differential metabolite production by the marine heterotroph, *Ruegeria pomeroyi*, in response to DMSP. ISME J. **10**: 2304–2316. doi:10.1038/ismej.2016.6

Abstract

Microbes, the foundation of the marine foodweb, do not function in isolation, but rather rely on molecular level interactions among species to thrive. While certain types of interactions between autotrophic and heterotrophic microorganisms have been well documented, the role of specific organic molecules in regulating inter-species relationships and supporting growth are only beginning to be understood. Here, we examine one such interaction by characterizing the metabolic response of a heterotrophic marine bacterium, *Ruegeria pomeroyi* DSS-3, to growth on dimethylsulfoniopropionate (DMSP), an abundant organosulfur metabolite produced by phytoplankton. When cultivated on DMSP, *R. pomeroyi* synthesized a quorum sensing molecule, *N*-(3-oxotetradecanoyl)-L-homoserine lactone, at significantly higher levels than during growth on propionate. Concomitant with the production of a quorum sensing molecule, we observed differential production of intra- and extracellular metabolites including glutamine, vitamin B₂, and biosynthetic intermediates of cyclic amino acids. Our metabolomics data indicate that *R. pomeroyi* changes regulation of its biochemical pathways in a manner that is adaptive for a cooperative lifestyle in the presence of DMSP, in anticipation of phytoplankton-derived nutrients and higher microbial density. This behavior is likely to occur on sinking marine particles, indicating that this response may impact the fate of organic matter.

3.1 Introduction

The carbon cycle is essential to the regulation of conditions on Earth from the climate to the pH buffering capacity of the ocean. Atmospheric carbon stores (750 Pg) are dwarfed by the 31,000 Pg of dissolved carbon present in the ocean, of which 660 Pg comprise dissolved organic matter. This complex pool of organic molecules contains substrates that serve as the primary catabolic energy source for heterotrophic bacteria, as well as compounds that are transferred among microbes to promote growth or cooperative community function. Because microbes have varying requirements for organic substrates, organic matter composition influences the microbial community and vice versa (Baña *et al.*, 2014; Sharma *et al.*, 2014). Some microbes are metabolically specialized while other microbes are generalists that can adapt to a variety of conditions (Lauro *et al.*, 2009). Recent studies have documented complex relationships between primary producers and heterotrophic bacteria (e.g. Durham *et al.*, 2015; Amin *et al.*, 2015). Experiments with the diatom *Thalassiosira pseudonana*, showed increased intracellular production of amino acids in response to co-culturing with the heterotroph *Dinoroseobacter shibae* (Paul *et al.*, 2012). In another study, the Roseobacter *Phaeobacter gallaeciensis* was found to produce phytotoxins in response to *p*-coumaric acid, an algal cell wall breakdown product (Seyedsayamdost *et al.*, 2011). More generally, vitamins such as cobalamin, biotin, and thiamin are produced by some heterotrophic bacteria and assimilated by eukaryotic organisms that are auxotrophic for these essential biochemicals (Croft *et al.*, 2006). Likewise, organosulfur metabolites are required by some marine bacteria, such as the ubiquitous SAR11, which cannot reduce inorganic sulfate (Tripp *et al.*, 2008). Finally, some molecules, including algal breakdown products and acyl homoserine lactones, act as infochemicals to signal events such as a nutrient pulse or population changes (Seyedsayamdost *et al.*, 2011; Sandy and Butler, 2009).

Although these interactions between microbes and molecules occur on very small spatial scales, they culminate in the production and remineralization of 50-100 Pg of carbon annually (Martin *et al.*, 1987).

DMSP is a widely recognized intermediate in these types of interactions. It is produced predominantly by dinoflagellates, coccolithophores, diatoms, and higher plants like the marsh grass *Spartina* (Stefels, 2000). Intracellularly, it is present at high concentrations ranging from 0.1-1 M (Stefels, 2000) and bulk particulate concentrations can reach 120 nM (Kiene and Slezak, 2006), making it a quantitatively significant organic substrate in the ocean (Simó *et al.*, 2009). Not surprisingly, it has been demonstrated to play a vital role in the marine food web as a source of reduced carbon and sulfur (González *et al.*, 2003; Simó *et al.*, 2009) and may act as an antioxidant in some organisms (Sunda *et al.*, 2002).

In the open ocean, dissolved DMSP concentrations are generally found in the sub-nanomolar range but can reach low nanomolar levels around a phytoplankton bloom and are likely elevated in the immediate wake of sinking aggregates of dying phytoplankton cells and/or fecal pellets (Kiene and Slezak, 2006; Levine *et al.*, 2012; Yoch, 2002). Consequently, DMSP is associated with the release of labile organic matter (Biddanda and Benner, 1997; Grossart, 2010; Stocker, 2012). Thus, heterotrophic organisms seeking hotspots of organic nutrients could reasonably use DMSP as a proxy for other, more complex, molecules. Indeed, previous work has shown that increased colonization occurs on DMSP-spiked agar particles relative to non-amended agar particles (Kjørboe *et al.*, 2002). Along these lines, DMSP acts as a chemo-attractant for organisms ranging from heterotrophic bacteria to herbivorous dinoflagellates (Seymour *et al.*, 2010). In one study, the presence of DMSP triggered the bacterial production of a plant growth hormone to promote growth of a phytoplankton species and thus augment organic

matter production (Seyedsayamdost *et al.*, 2011).

Bacteria with relatively large genomes can take advantage of transient nutrient pulses and thus could be well-suited to respond to DMSP-derived signals. To examine the metabolic response of a heterotrophic bacterium to DMSP, we focus here on the metabolically-flexible bacterium, *Ruegeria pomeroyi* DSS-3, a coastal isolate of the Roseobacter clade of marine α -proteobacteria (Moran *et al.*, 2004). *R. pomeroyi* possesses genes for a variety of metabolic strategies including lithoheterotrophy using carbon monoxide or sulfide, flagellar motility, and transporters for algal osmolytes and carboxylic acids (Moran *et al.*, 2004). These attributes, which have been confirmed in culture, enable *R. pomeroyi* to exploit rapid chemical changes in its environment. In particular, it can catabolize DMSP via two different pathways and incorporate atoms derived from DMSP into its biomass (González *et al.*, 2003; Reisch *et al.*, 2013). There are also indications that DMSP may provide a signal to prime *R. pomeroyi*'s metabolism to utilize algal lysates and exudates through increased transcription of transporter genes as well as a potential quorum sensing response (Bürgmann *et al.*, 2007). Thus, *R. pomeroyi* is likely to have the metabolic capability to respond to DMSP as an indicator of a shift in nutrient availability.

We examined *R. pomeroyi*'s relationship with DMSP by comparing its metabolic response to DMSP as a carbon source relative to propionate, a 3-carbon carboxylic acid that lacks the dimethylsulfide functional group. If DMSP serves only as an organic substrate for *R. pomeroyi*, then the primary metabolic difference in our experiment should be a response to the reduced sulfur in DMSP (Reisch *et al.*, 2013) because propionate catabolism shares the last two steps of one of the DMSP catabolic pathways. Alternatively, if DMSP indeed functions as a signal to *R. pomeroyi* of changing environmental conditions, we would anticipate a suite of

metabolic changes that facilitate adaptation to a DMSP-containing microzone.

The microzones in the ocean where DMSP is prevalent have several characteristics in common. Specifically, these microzones have higher nutrient availability compared to the surrounding water column, particularly organic substrates which can provide nitrogen and phosphorus, and they host a dense microbial community whose members compete for, and exchange, valuable substrates and nutrients (Grossart, 2010; Biddanda and Benner, 1997). In response to these characteristics, *R. pomeroyi* and similar metabolically-flexible bacteria could change their metabolism to take advantage of the shift in chemical regime. Here we examine three hypotheses using intracellular and extracellular metabolic profiles. First, higher cell densities facilitate cell-cell communication (or quorum sensing) and resultant coordinated behaviors, and thus we expect to resolve and identify chemical signaling molecules. Second, the most energetically favorable means of obtaining nutrients, particularly nitrogen, would be to acquire them directly from the labile organic matter ubiquitous in the microzones (Kirchman *et al.*, 1989, 2003; Vallino *et al.*, 1996). Bürgmann *et al.* (2007) demonstrated that *R. pomeroyi* increases transcription of genes associated with transporters for nitrogen-containing organic substrates when grown on DMSP. We anticipate that a change in nutrient acquisition strategy would also result in differences in the concentrations of regulatory metabolites. Last, we suggest that a mutually beneficial exchange of high-value metabolites with other bacteria or phytoplankton would occur (Amin *et al.*, 2015; Durham *et al.*, 2015), leading to an adaptive and cooperative lifestyle similar to that observed during bacterial colonization of organic particles (Grossart *et al.*, 2003).

Our study examines the impact of DMSP relative to propionate on metabolite production and organic matter excretion by *R. pomeroyi*. We used complementary untargeted and targeted

metabolomics methods to analyze the suite of molecules produced by *R. pomeroyi*, both within the cells and in the external medium. These metabolites ranged from those associated with primary metabolism, which shed light on fundamental cellular processes, to those associated with secondary metabolism, which help us understand the cell's response to shifting environmental conditions.

3.2 Materials and Methods

The complete methods are given in the supplemental information. Briefly, *Ruegeria pomeroyi* DSS-3 [DSM 15171] was grown in the dark at 23°C in 0.2 µm-filtered (Omnipore (PTFE), EMD Millipore) seawater amended with 3 mM carbon (1 mM sodium propionate or 0.6 mM DMSP), 4 mM ammonium chloride, 30 nM monosodium phosphate, 100 nM ferrous chloride-ethylenediaminetetraacetic acid, 100 nM zinc chloride, 100 nM manganese(II) chloride, 1 nM cobalt(II) chloride, and 1 mL / L medium of f/2 vitamin solution. Duplicate cultures were sacrificed at 0, 32, 38(propionate) /43(DMSP), 48, 60, and 72 hours with cell counts in the range of 10^6 - 10^7 cells/mL (Fig. S1). The third time point was offset between the propionate and DMSP treatments so that the same cell abundance was sampled in each treatment. Cells were captured on a 0.2 µm filter (Omnipore (PTFE), EMD Millipore) using a combusted glass vacuum filtration apparatus (vacuum never greater than 10 mmHg). Filters were frozen in cryovials at -80°C immediately after filtration. Intracellular metabolites were extracted from filters using cold 40:40:20 acetonitrile:methanol:water + 0.1 M formic acid solution, adapted from previously published protocols (Rabinowitz and Kimball, 2007). Half of the extract was further extracted using Agilent Bond Elut PPL (styrene-divinylbenzene polymer) solid-phase extraction resin for untargeted analysis. The extracellular metabolites were extracted from acidified filtrate (pH 2-3)

using PPL cartridges (Dittmar *et al.*, 2008; after modification by Longnecker, 2015).

All samples were separated with the same liquid chromatography method on a C18 reversed phase column with a water and acetonitrile gradient (Kido Soule *et al.*, 2015). The targeted method used a Thermo Scientific triple quadrupole mass spectrometer with a heated electrospray ionization source. Untargeted analyses were conducted using a hybrid linear ion trap - 7T Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR-MS; LTQ FT Ultra, Thermo Scientific) with an electrospray ionization source.

The targeted data were processed using Xcalibur (Thermo Scientific software) and a 5 to 7-point standard curve for relative quantification. The untargeted mass analysis files were converted from Thermo RAW files to mzML files using MSConvert (Chambers *et al.*, 2012). The resulting data were processed using XCMS (Smith *et al.*, 2006; Tautenhahn *et al.*, 2008; Benton *et al.*, 2010) and CAMERA (Kuhl *et al.*, 2012). Peak picking and alignment were carried out separately for the intracellular and extracellular samples generating a list of features (defined as a unique combination of mass-to-charge value (m/z) and retention time) for each sample set (Longnecker *et al.*, 2015). Feature lists were refined through quality control requirements and are reported as mass spectral peak area normalized to cell abundance (intra- and extracellular; peak area / cell) or volume (extracellular; peak area / L). Features in the untargeted data were putatively identified by comparison of m/z to the METLIN database (Smith *et al.*, 2005), and if available, by comparison of the associated fragmentation spectrum with experimentally-generated fragmentation patterns in the METLIN database or to *in silico* derived fragmentation patterns from MetFrag (Wolf *et al.*, 2010). When possible, commercially available analogs for putatively identified metabolites were analyzed using identical protocols.

Differences in the concentration of individual metabolites between the substrate

treatments were compared with ANOVA. In the intracellular targeted data, p-values < 0.0013 were considered significant due to the Bonferroni correction for multiple comparisons.

Total organic carbon (TOC; unfiltered) and dissolved organic carbon (DOC; 0.2- μ m filtered) samples were acidified and then analyzed using a Shimadzu TOC-V_{CSH} Total Organic Carbon Analyzer, according to standard practices. Methanethiol and dimethylsulfide were measured using gas chromatography (adapted from Levine *et al.*, 2012).

3.3 Results

3.3.1 *R. pomeroyi* Growth Parameters

R. pomeroyi cells exhibited similar growth rates (0.1 h⁻¹) on both substrates (DMSP or propionate), and remineralization of the organic carbon substrate was confirmed by the diminishing concentrations of DOC and TOC in the media (Fig. S1). Methanethiol (MeSH) and dimethylsulfide (DMS), which are products of two different DMSP degradation pathways, were produced during growth on DMSP (Fig. S1; Reisch *et al.*, 2013, 2011). In this study we did not quantify the two gases as we only sought confirmation that both pathways were active.

3.3.2 Quantifying the metabolome (Targeted Metabolites)

In targeted metabolomics, the goal is to obtain relative concentrations of a pre-defined list of organic compounds. Most of the intracellular and extracellular metabolites quantified by our method showed no significant difference between the substrate treatments. For example, intracellular methionine concentrations were similar in both treatments (Fig. 1d). The measured compounds in the targeted method include vitamins, amino acids, nucleotides, and precursors to these molecules, implying that primary cellular processes remained similar regardless of substrate (see Fig. S2 for the list of intracellular metabolites quantified). However, within the

targeted data encompassing both intra and extracellular samples, significant changes were observed for one of the amino acids (glutamine), several vitamins, and intermediates in cellular nitrogen and sulfur cycling. Intracellular glutamine concentrations by time point ranged from 3.5-157 times lower during growth on DMSP relative to growth on propionate (Fig. 1a). This varied by time in the growth curve and the concentrations were particularly low in the DMSP treatment. Conversely, the B vitamins thiamin monophosphate (phosphorylated vitamin B₁) and riboflavin (vitamin B₂) were both elevated within the cell at some time points in the DMSP treatment (Fig. 1b and 1c, respectively). Extracellularly, the maximum concentration of riboflavin was almost eight times higher in the DMSP treatment (Fig. 2a).

The intracellular concentrations of 5'-deoxy-5'-methylthioadenosine (MTA), a sulfur-containing metabolite, were not significantly different between the two treatments (Fig. 3b). However, in the extracellular fraction, MTA concentrations were approximately three times higher in the DMSP treatment than in the propionate treatment (Fig. 3c), indicating increased production and release of MTA into the medium during DMSP-based growth. Given its sulfur content, the elevated MTA concentration could be derived from excess sulfur generated during DMSP catabolism. We tested this hypothesis with an experiment that monitored MTA concentrations in three treatments: 100% propionate, 90% propionate / 10% DMSP and 100% DMSP, where all treatments contained the same amount of dissolved organic carbon (3 mM). When normalized to cell abundance, the extracellular MTA concentrations in the full DMSP and the 10% DMSP treatments were seven and six times higher, respectively, than in the propionate treatment (Fig. S3a). Therefore, despite the 90% reduction in sulfur, there was not a corresponding decrease in cell-normalized MTA concentration.

3.3.3 Profiling the Metabolome (Untargeted Metabolites)

The goal of the untargeted method is to profile as much of the metabolome as possible. Unlike the targeted method it is only semi-quantitative and relative abundances cannot be converted into molar quantities. However, metabolites measured in both the targeted and untargeted methods show highly similar instrument responses between the two methods. This is shown here for MTA (Fig. S3b) and previous work has demonstrated this for other metabolites (Fiore *et al.*, 2015). This confirms that relative abundances of features (i.e., molecules defined by a unique combination of m/z value and retention time) within the untargeted dataset can be compared and statistically evaluated.

Hundreds of intracellular features were resolved with the untargeted metabolomics method (Table S1). Most of these (43%, negative ion mode; 96%, positive ion mode) were detected in both substrate treatments but their relative abundances illustrate more striking differences between the treatments (Table S2). At 60 hours, a majority (86%, negative ion mode; 64%, positive ion mode) of the intracellular features were 5 times higher in the propionate treatment. In contrast, a minority (9%, negative ion mode; 17%, positive ion mode) of the intracellular features were at least five times higher in the DMSP treatment relative to the propionate treatment (Table S2).

Hundreds of extracellular features were similarly observed in these samples (Table S1). A non-metric multidimensional scaling analysis of the extracellular metabolic profiles shows that samples cluster more by cell density and/or growth phase than by substrate treatment (Fig. S4). Carbon substrate is of secondary importance but has a discernible impact, particularly at later time points. A comparison of peak area/cell of the individual extracellular features at 60 hours shows that, depending on substrate and ionization mode, between 19-36% of these features were at least five times higher in one of the substrate treatments (Table S2).

3.3.4 Identification of Features in the Untargeted Dataset

The structural diversity of metabolites and the gaps in the characterization of metabolic pathways present major challenges to the identification of features in the untargeted dataset. Our approach was to identify as many features as possible within the ranking system proposed by Sumner et al. (2007) and described in Longnecker et al. (2015). Metabolites are identified with increasing confidence first by an m/z match, then a fragmentation match to a database, and, for the highest confidence level, comparison to an authentic standard analyzed with our method to confirm m/z value, retention time, and fragmentation pattern. In some cases, putative identifications of metabolites with only an m/z match can be considered more significant if intermediates in the same metabolic pathway are identified at higher rankings (Suhre and Schmitt-Kopplin, 2008). These stringent requirements result in a relatively small subset of metabolites that we can reliably identify. However, we can expect to be able to identify a larger set of metabolites in future studies as the field of metabolomics matures; these additional identifications will add greater nuance to our understanding of metabolic responses and their ecological implications.

We focused our most rigorous identification efforts, including purchasing standards, on features that were differentially produced in the two treatments. MTA was quantified by our targeted method and found to be elevated in the extracellular fraction of the DMSP growth treatment; it was similarly detected in the untargeted method (Fig. S3b). MTA can be produced by two different pathways, as a by-product of spermidine / spermine synthesis or acyl homoserine lactone (AHL) biosynthesis (Parveen and Cornell, 2011) (Fig. S5). Structurally, AHLs consist of a lactone head-group linked by an amide to a carbon chain of varying lengths and degrees of saturation (Fig. 3d, inset). AHLs are a class of quorum-sensing chemicals and *R.*

pomeroyi is known to produce three different AHL compounds (Moran *et al.*, 2004). We searched the untargeted dataset for predicted m/z values for all AHLs up to a 16-carbon chain length. We identified *N*-(3-oxotetradecanoyl)-L-homoserine lactone in our dataset, based on exact mass and fragmentation pattern matches in the METLIN database (Smith *et al.*, 2005). Its size was consistent with an AHL observed by Moran *et al.* (2004). Comparison with an authentic standard revealed a positive identification for this compound (Fig. 4). The identified AHL is elevated in the extracellular medium by a factor of 8-15 in the DMSP treatment at the last four time points of the experiment (Fig. 3d), where MTA concentrations are also high.

We also putatively identified precursors to the cyclic amino acids (tryptophan, phenylalanine, and tyrosine) including chorismate, shikimate, phenylpyruvate, and their associated intermediates (Table S3; Fig. S6). These features occur at higher relative concentrations in the DMSP treatment (Fig. 2). Other putative identifications include the first metabolite in the DMSP demethylation pathway, 3-(methylthio)propionate, which was exclusively observed in the DMSP treatment in both extracellular and intracellular fractions (Fig. 2; Table S3; Table S4). Additionally, in the untargeted data, α -ribazole, a component of vitamin B₁₂, was putatively identified and found to accumulate in the media over time (Fig. 2b; Table S3). While the current annotations of *R. pomeroyi*'s genome do not predict direct synthesis of α -ribazole, its production may result from an un-annotated biosynthetic pathway or non-enzymatic degradation of vitamin B₁₂.

3.4 Discussion

The primary goal of this study was to examine the difference between the two growth substrates (i.e., DMSP and propionate) beyond the obvious differences in direct degradation of

the carbon substrate. We hypothesized that these differences would be observed either in sulfur-containing metabolites (Thiel *et al.*, 2010) or in metabolites related to a re-tooling of metabolism for a more cooperative lifestyle (Geng *et al.*, 2008; Seyedsayamdost *et al.*, 2011). While a few sulfur-containing metabolites, such as MTA, had a higher abundance in the DMSP treatment, we did not find any indication of far-reaching effects on sulfur-containing metabolites under DMSP-growth. For instance, an essential metabolite such as methionine, a sulfur-containing amino acid, had equivalent concentrations in the two treatments (Fig. 1d) suggesting its intracellular concentration is highly regulated and insensitive to the influx of sulfur. Thus, rather than an impact on only S-containing molecules, we observed a broad shift in metabolite patterns, suggesting a pervasive impact of DMSP on the *R. pomeroyi* metabolome.

3.4.1 DMSP Degradation

The fate of DMSP during degradation by *R. pomeroyi* has been extensively studied (Reisch *et al.*, 2013) and the concentrations of metabolites within the DMSP degradation pathways are consistent with previous genomics research. Based on the metabolites we measured, both the DMSP demethylation and the cleavage pathways were active in the DMSP treatment (Fig. 2; Table S4; Fig. S1). After the production of these molecules, the atoms derived from DMSP are incorporated into intermediates within central carbon metabolism and are impossible to discern relative to propionate metabolism (without isotope labeling).

3.4.2 MTA as a proxy for possible quorum sensing in *R. pomeroyi*

MTA, which accumulates in the DMSP treatment, is an intermediate in the methionine salvage pathway, which allows the cell to recycle sulfur atoms for methionine synthesis. *R. pomeroyi* lacks the genes that complete the classic methionine salvage pathway, suggesting that excess MTA produced by *R. pomeroyi* would be excreted rather than recycled. MTA is a by-

product of AHL synthesis and of spermidine/spermine synthesis (Fig. S5) (Parveen and Cornell, 2011). *R. pomeroyi*'s genome suggests that it can synthesize spermidine and spermine, but that this occurs through an alternative pathway that does not produce MTA. Therefore, of the characterized pathways that generate MTA as a by-product, the *R. pomeroyi* genome is only known to contain genes required for AHL synthesis (Fig. S5). Furthermore, Bürgmann et al. (2007) have demonstrated increased transcription of a gene annotated as a “transcriptional regulator, LuxR family” (*luxR* is the AHL receptor gene) in response to DMSP. Although *luxR* and *luxI* (the AHL synthase gene) did not show increased transcription in the Bürgmann et al. (2007) study, the regulation of these systems is complex, as shown in species closely related to *R. pomeroyi* (Zan et al., 2014), and might not be visible in the transcriptome for a variety of reasons. Based on this genomic information, MTA was most likely found at higher concentrations in the DMSP treatment due to increased AHL synthesis. This hypothesis was supported in the untargeted data by the detection of one AHL, *N*-(3-oxotetradecanoyl)-L-homoserine lactone. The coherence of MTA and AHL dynamics suggests that the elevated MTA concentrations are linked to increased synthesis of *N*-(3-oxotetradecanoyl)-L-homoserine lactone and thus the presence of a possible quorum sensing response by *R. pomeroyi* to elevated DMSP concentrations. A linear decrease in MTA concentration was not observed between a culture grown on 100% DMSP and one grown on 10% DMSP/90% propionate, further implying that the elevated accumulation of MTA in the DMSP treatment is not simply due to the increased reduced sulfur pool, but instead is the result of a regulatory response triggered by DMSP.

Phenotypes regulated by a quorum sensing system generally depend on a high density of cells to be effective. Quorum sensing can regulate activities such as antibiotic production, exoenzyme synthesis, or biofilm formation (Miller and Bassler, 2001). *R. pomeroyi* can produce

exoenzymes to break down large biopolymers (Christie-Oleza *et al.*, 2015) and has genes for flagellar motility which can be regulated by a quorum sensing system (Moran *et al.*, 2004). The microzones with likely elevated DMSP concentrations generally have high microbial population densities. With a population-dependent cell-cell communication pathway, such as AHL-based quorum sensing, *R. pomeroyi* could coordinate a shift in its metabolism to exploit the increase in organic nutrient availability. This hypothesis is consistent with previous characterization of a quorum-sensing phenotype of *Ruegeria* sp. KLH11, a marine sponge symbiont (Zan *et al.*, 2012), and with recent findings of measurable quantities of a number of AHLs in association with sinking marine particles (Hmelo *et al.*, 2011). Further work will be needed to understand the specificity of *R. pomeroyi*'s response to DMSP including experiments that evaluate *R. pomeroyi* metabolic (and transcriptomic / proteomic) profiles under exposure to the AHL identified here, as well as experiments with *luxR/luxI* knock-out mutants to isolate the affected metabolic pathways.

3.4.3 Glutamine as a proxy for altered nitrogen assimilation strategy

It is particularly striking that intracellular glutamine concentrations were relatively reduced in the DMSP treatment because it is a metabolite that is highly sensitive to nitrogen availability and is integral to a cell's nitrogen cycle (Brauer *et al.*, 2006). Glutamine can serve as a nitrogen storage compound because it is produced when ammonium is assimilated into the cell. Once present, glutamine is an essential precursor for the synthesis of other amino acids and donates an amino group during the synthesis of purines, pyrimidines, amino sugars, and NAD⁺ (Reitzer, 2003). High glutamine concentrations indicate that sufficient ammonium concentrations are present. In contrast, low glutamine concentrations signal a decline in ammonium concentrations and result in the up-regulation of alternative nitrogen assimilation pathways. This

has been observed not only in *E. coli*, but also in organisms such as *Saccharomyces cerevisiae* (Brauer *et al.*, 2006). In our experiments, excess ammonium was present in the growth media in both treatments, and thus inorganic nitrogen availability can be discounted as a controlling factor of intracellular glutamine concentrations.

Through a cascade of signals, glutamine also regulates alternative forms of nitrogen acquisition. In *E. coli*, for example, low glutamine concentrations are correlated with increased transport of organic nitrogen substrates such as amino acids, peptides, and polyamines into the cell to offset the decline in extracellular ammonium (Zimmer *et al.*, 2000; Chubukov *et al.*, 2014). While the regulation of nitrogen assimilation has not been studied in *R. pomeroyi*, *R. pomeroyi*'s genome annotations indicate the presence of genes for both pathways for ammonium assimilation (glutamate dehydrogenase and glutamine synthetase) as well as for the nitrogen regulatory protein P-II and a uridylyltransferase, all of which would be required to regulate nitrogen assimilation in a manner similar to *E. coli*. Bürgmann *et al.* (2007) previously observed that *R. pomeroyi*'s transcriptional response to DMSP involved up-regulation of genes specifically associated with amino acid and polyamine transport. This accumulated evidence suggests that *R. pomeroyi* altered its nitrogen assimilation strategy in response to the presence of DMSP. Specifically, we hypothesize that *R. pomeroyi* is regulating intracellular glutamine concentrations in order to promote organic nitrogen uptake and utilization. Scavenging nitrogen from organic sources rather than utilizing ammonium could benefit *R. pomeroyi* by decreasing the cellular energy expenditure associated with ammonium assimilation and/or full biosynthesis of nitrogenous metabolites. To test this further, experiments are needed that compare the uptake of organic nitrogen sources and the differential transcription of nitrogen assimilation genes, in both the presence and absence of DMSP.

3.4.4 Production of extracellular metabolites with potential for microbial cross-feeding

Reduced glutamine concentrations can also affect the diverse biosynthetic pathways that rely on this metabolite. For example, during nitrogen starvation in *E. coli*, decreased glutamine concentrations result in lower tryptophan concentrations and accumulation of phenylpyruvate, a precursor of phenylalanine, due to decreased *de novo* synthesis of phenylalanine from phenylpyruvate (Brauer *et al.*, 2006). Tryptophan shares a common precursor, chorismate, with the other cyclic metabolites phenylalanine, tyrosine, 2,3-dihydroxybenzoic acid, and 4-aminobenzoic acid (Fig. S6). In our experiment, intracellular and extracellular tryptophan concentrations remain the same in both treatments, while its precursors, 3-dehydroshikimate, shikimate, chorismate, and phenylpyruvate, all accumulate in the extracellular fraction of the DMSP treatment. This suggests that the precursors for tryptophan synthesis were not being utilized by *R. pomeroyi* either due to decreased synthesis of the amino acid or to increased production of the precursors.

The cellular release of these metabolically high-value biosynthetic intermediates supports the third component of our hypothesis, namely that switching to a cooperative lifestyle triggered by DMSP would require the release of some metabolites that could be used in cross-feeding. The increased extracellular concentrations of precursors like chorismate and phenylpyruvate could provide labile intermediates to surrounding microbes with incomplete amino acid biosynthetic pathways; these organisms might in turn produce valuable metabolites that *R. pomeroyi* could use. According to the annotated *R. pomeroyi* genome, *R. pomeroyi* can synthesize all amino acids except asparagine and possibly histidine. Analysis of 3062 available bacterial genomes in the Integrated Microbial Genomes (IMG) database reveals that biosynthetic capability in bacteria ranges from 0 to 20 amino acids, averaging only 7.9 complete amino acid biosynthetic pathways.

Bacteria are most commonly auxotrophic for tyrosine, phenylalanine, lysine, or histidine. In 124 eukaryote genomes, complete biosynthetic pathways are present, on average, for only 4.1 out of 20 amino acids with auxotrophy most commonly found for histidine, lysine, serine, or leucine and no species able to synthesize more than 14 of the 20 essential amino acids (Mee and Wang, 2012). The enhanced excretion of these biosynthetically high-value cyclic intermediates under DMSP-derived growth may be linked in part to changes in nitrogen assimilation strategy used by *R. pomeroyi*. However, these metabolites would also serve as valuable currency for microbial cross-feeding as evidenced by the ubiquity of incomplete *de novo* amino acid synthesis pathways.

Other metabolites that could be used in microbial cross-feeding are riboflavin (vitamin B₂) and α -ribazole, a component of vitamin B₁₂. Both of these molecules accumulated in the extracellular matrix of the DMSP treatment (riboflavin to 8X and α -ribazole to 3X at 72hrs; Fig. 2). It is reasonable to speculate that the availability of a precursor such as α -ribazole, as well as the vitamin, riboflavin, would be valuable within a dense microbial consortium. For example, auxotrophy for α -ribazole has been demonstrated in the bacterium, *Listeria innocua*, which has a specific transporter to acquire this compound from its environment (Gray and Escalante-Semerena, 2010). Although α -ribazole, as well as the cyclic amino acid precursors mentioned earlier, do not directly provide amino acids and vitamins, they would allow organisms to synthesize essential metabolites at lower metabolic cost and could lead to precursor auxotrophy (the “black queen” hypothesis (Morris *et al.*, 2012)). Further analysis of marine genomes may improve our ability to link metabolite production to the demands of different types of auxotrophic organisms but clearly exchange of metabolites, as proposed here, would be mutually beneficial to microbes in a high-density consortium.

3.5 Conclusions

Differences in the phylogenetic diversity of free-living versus surface-associated microbial populations have been well-documented (DeLong *et al.*, 1993). However, many bacteria may switch between these two lifestyles and the factors governing this switch are only beginning to be understood (Grossart, 2010). Our work expands on these studies by showing that specific metabolites, such as algal-derived DMSP, can trigger metabolic shifts that could be expressed under high population densities (Fig. 5). If groups of bacteria are using some phytoplankton-derived metabolites as signals (Seyedsayamdost *et al.*, 2011), rates of particle degradation, and more broadly of carbon remineralization, may be intimately tied to particle composition and source. These results emphasize the need to understand the molecular-level composition of organic matter in the ocean because a single molecule, such as DMSP, can alter the metabolism of organisms that encounter it, with likely impact on their biogeochemical function.

Acknowledgments

This research is funded in part by the Gordon and Betty Moore Foundation through Grant GBMF3304 as well as by the National Science Foundation (Grants OCE-0928424 and OCE-1154320). The instruments in the WHOI FT-MS Facility were purchased with support from the GBMF and NSF. Support for WMJ came from a National Defense Science and Engineering Fellowship. In addition the authors thank the following people for their assistance: Krista Longnecker and Cara Fiore for help with data analysis, Cara Fiore and Isa Howard-Åkerfeldt for technical assistance, Jamey Fulton and Naomi Levine for assistance with the GC-PFPD,

Bethanie Edwards for instrument instruction, and Ben Van Mooy, Mary Ann Moran, and Phil Gschwend for feedback on this work. The raw mass spectrometry files and processed peak files have been archived at MetaboLights (MTBLS157, Release date: 30 January 2015) run by the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI).

References

- Amin SA, Hmelo LR, van Tol HM, Durham BP, Carlson LT, Heal KR, *et al.* (2015). Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **522**:98–101.
- Baña Z, Ayo B, Marrasé C, Gasol JM, Iriberry J. (2014). Changes in bacterial metabolism as a response to dissolved organic matter modification during protozoan grazing in coastal Cantabrian and Mediterranean waters. *Environ Microbiol* **16**:498–511.
- Benton HP, Want EJ, Ebbels TMD. (2010). Correction of mass calibration gaps in liquid chromatography-mass spectrometry metabolomics data. *Bioinformatics* **26**:2488–2489.
- Biddanda B, Benner R. (1997). Carbon, nitrogen, and carbohydrate fluxes during the production of particulate and dissolved organic matter by marine phytoplankton. *Limnol Oceanogr* **42**:506–518.
- Brauer MJ, Yuan J, Bennett BD, Lu W, Kimball E, Botstein D, *et al.* (2006). Conservation of the metabolomic response to starvation across two divergent microbes. *Proc Natl Acad Sci U S A* **103**:19302–19307.
- Bürgmann H, Howard EC, Ye W, Sun F, Sun S, Napierala S, *et al.* (2007). Transcriptional response of *Silicibacter pomeroyi* DSS-3 to dimethylsulfoniopropionate (DMSP). *Environ Microbiol* **9**:2742–2755.
- Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, *et al.* (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **30**:918–920.
- Christie-Oleza JA, Scanlan DJ, Armengaud J. (2015). ‘You produce while I clean up,’ a strategy revealed by exoproteomics during *Synechococcus-Roseobacter* interactions. *Proteomics* **15**:3454–3462.
- Chubukov V, Gerosa L, Kochanowski K, Sauer U. (2014). Coordination of microbial metabolism. *Nat Rev Microbiol* **12**:327–340.
- Croft MT, Warren MJ, Smith AG. (2006). Algae need their vitamins. *Eukaryot Cell* **5**:1175–1183.
- DeLong EF, Franks DG, Alldredge AL. (1993). Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. *Limnol Oceanogr* **38**:924–934.
- Dittmar T, Koch B, Hertkorn N, Kattner G. (2008). A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol Oceanogr Methods* **6**:230–235.
- Durham BP, Sharma S, Luo H, Smith CB, Amin SA, Bender SJ, *et al.* (2015). Cryptic carbon and sulfur cycling between surface ocean plankton. *Proc Natl Acad Sci U S A* **112**:453–457.
- Fiore CL, Longnecker K, Kido Soule MC, Kujawinski EB. (2015). Release of ecologically relevant metabolites by the cyanobacterium, *Synechococcus elongatus* CCMP 1631. *Environ Microbiol* **17**:3949–3963.
- Geng H, Bruhn JB, Nielsen KF, Gram L, Belas R. (2008). Genetic dissection of tropodithietic acid biosynthesis by marine roseobacters. *Appl Environ Microbiol* **74**:1535–1545.
- González JM, Covert JS, Whitman WB, Henriksen JR, Mayer F, Scharf B, *et al.* (2003).

- Silicibacter pomeroyi* sp. nov. and *Roseovarius nubinhibens* sp. nov., dimethylsulfoniopropionate-demethylating bacteria from marine environments. *Int J Syst Evol Microbiol* **53**:1261–1269.
- Gray MJ, Escalante-Semerena JC. (2010). A new pathway for the synthesis of α -ribazole-phosphate in *Listeria innocua*. *Mol Microbiol* **77**:1429–1438.
- Grossart H, Kjørboe T, Tang K, Ploug H. (2003). Bacterial colonization of particles: growth and interactions. *Appl Environ Microbiol* **69**:3500–3509.
- Grossart H-P. (2010). Ecological consequences of bacterioplankton lifestyles: changes in concepts are needed. *Environ Microbiol Rep* **2**:706–714.
- Hmelo LR, Mincer TJ, Van Mooy BAS. (2011). Possible influence of bacterial quorum sensing on the hydrolysis of sinking particulate organic carbon in marine environments. *Environ Microbiol Rep* **3**:682–688.
- Kido Soule MC, Longnecker K, Johnson WM, Kujawinski EB. (2015). Environmental metabolomics: analytical strategies. *Mar Chem* **177**:374–387.
- Kiene RP, Slezak D. (2006). Low dissolved DMSP concentrations in seawater revealed by small volume gravity filtration and dialysis sampling. *Limnol Oceanogr Methods* **4**:80–95.
- Kjørboe T, Grossart H-P, Ploug H, Tang K. (2002). Mechanisms and rates of bacterial colonization of sinking aggregates. *Appl Environ Microbiol* **68**:3996–4006.
- Kirchman DL, Hoffman KA, Weaver R, Hutchins DA. (2003). Regulation of growth and energetics of a marine bacterium by nitrogen source and iron availability. *Mar Ecol Prog Ser* **250**:291–296.
- Kirchman DL, Keil RG, Wheeler PA. (1989). The effect of amino acids on ammonium utilization and regeneration by heterotrophic bacteria in the subarctic Pacific. *Deep Sea Res Part A Oceanogr Res Pap* **36**:1763–1776.
- Kuhl C, Tautenhahn R, Bottcher C, Larson TR, Neumann S. (2012). CAMERA: An integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal Chem* **84**:283–289.
- Lauro FM, McDougald D, Thomas T, Williams TJ, Egan S, Rice S, *et al.* (2009). The genomic basis of trophic strategy in marine bacteria. *Proc Natl Acad Sci U S A* **106**:15527–15533.
- Levine NM, Varaljay VA, Toole DA, Dacey JWH, Doney SC, Moran MA. (2012). Environmental, biochemical and genetic drivers of DMSP degradation and DMS production in the Sargasso Sea. *Environ Microbiol* **14**:1210–1223.
- Longnecker K. (2015). Dissolved organic matter in newly formed sea ice and surface seawater. *Geochim Cosmochim Acta* **171**:39–49.
- Longnecker K, Futrelle J, Coburn E, Kido Soule MC, Kujawinski EB. (2015). Environmental metabolomics: Databases and tools for data analysis. *Mar Chem* **177**:366–373.
- Martin J, Knauer G, Karl D, Broenkow W. (1987). VERTEX: Carbon cycling in the northeast Pacific. *Deep Sea Res Part A Oceanogr Res Pap* **34**:267–285.
- Mee MT, Wang HH. (2012). Engineering ecosystems and synthetic ecologies. *Mol Biosyst* **8**:2470–2483.

- Miller M, Bassler B. (2001). Quorum sensing in bacteria. *Annu Rev Microbiol* **55**:165–199.
- Moran MA, Buchan A, González JM, Heidelberg JF, Whitman WB, Kiene RP, *et al.* (2004). Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature* **432**:910–913.
- Morris JJ, Lenski RE, Zinser ER. (2012). The black queen hypothesis: Evolution of dependencies through adaptive gene loss. *mBio* **3**:1–7.
- Parveen N, Cornell KA. (2011). Methylthioadenosine/S-adenosylhomocysteine nucleosidase, a critical enzyme for bacterial metabolism. *Mol Microbiol* **79**:7–20.
- Paul C, Mausz MA, Pohnert G. (2012). A co-culturing/metabolomics approach to investigate chemically mediated interactions of planktonic organisms reveals influence of bacteria on diatom metabolism. *Metabolomics* **9**:349–359.
- Rabinowitz JD, Kimball E. (2007). Acidic acetonitrile for cellular metabolome extraction from *Escherichia coli*. *Anal Chem* **79**:6167–6173.
- Reisch CR, Crabb WM, Gifford SM, Teng Q, Stoudemayer MJ, Moran MA, *et al.* (2013). Metabolism of dimethylsulphoniopropionate by *Ruegeria pomeroyi* DSS-3. *Mol Microbiol* **89**:774–791.
- Reisch CR, Moran MA, Whitman WB. (2011). Bacterial catabolism of dimethylsulfoniopropionate (DMSP). *Front Microbiol* **2**:172–184.
- Reitzer L. (2003). Nitrogen assimilation and global regulation in *Escherichia coli*. *Annu Rev Microbiol* **57**:155–176.
- Sandy M, Butler A. (2009). Microbial iron acquisition: marine and terrestrial siderophores. *Chem Rev* **109**:4580–4595.
- Seyedsayamdost MR, Case RJ, Kolter R, Clardy J. (2011). The Jekyll-and-Hyde chemistry of *Phaeobacter gallaeciensis*. *Nat Chem* **3**:331–335.
- Seymour J, Simó R, Ahmed T, Stocker R. (2010). Chemoattraction to dimethylsulfoniopropionate throughout the marine microbial food web. *Science (80-)* **329**:342–345.
- Sharma AK, Becker JW, Ottesen EA, Bryant JA, Duhamel S, Karl DM, *et al.* (2014). Distinct dissolved organic matter sources induce rapid transcriptional responses in coexisting populations of *Prochlorococcus*, *Pelagibacter* and the OM60 clade. *Environ Microbiol* **16**:2815–2830.
- Simó R, Vila-Costa M, Alonso-Sáez L, Cardelús C, Guadayol Ò, Vázquez-Domínguez E, *et al.* (2009). Annual DMSP contribution to S and C fluxes through phytoplankton and bacterioplankton in a NW Mediterranean coastal site. *Aquat Microb Ecol* **57**:43–55.
- Smith CA, Maille GO, Want EJ, Qin C, Trauger SA, Brandon TR, *et al.* (2005). A metabolite mass spectral database. *Ther Drug Monit* **27**:747–751.
- Smith CA, Want EJ, O’Maille G, Abagyan R, Siuzdak G. (2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **78**:779–787.
- Stefels J. (2000). Physiological aspects of the production and conversion of DMSP in marine algae and higher plants. *J Sea Res* **43**:183–197.

- Stocker R. (2012). Marine Microbes See a Sea of Gradients. *Science* (80-) **338**:628–633.
- Suhre K, Schmitt-Kopplin P. (2008). MassTRIX: mass translator into pathways. *Nucleic Acids Res* **36**:481–484.
- Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, *et al.* (2007). Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3**:211–221.
- Sunda W, Kieber DJ, Kiene RP, Huntsman S. (2002). An antioxidant function for DMSP and DMS in marine algae. *Nature* **418**:317–320.
- Tautenhahn R, Böttcher C, Neumann S. (2008). Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* **9**:504–520.
- Thiel V, Brinkhoff T, Dickschat JS, Wickel S, Grunenberg J, Wagner-Döbler I, *et al.* (2010). Identification and biosynthesis of tropone derivatives and sulfur volatiles produced by bacteria of the marine Roseobacter clade. *Org Biomol Chem* **8**:234–246.
- Tripp HJ, Kitner JB, Schwalbach MS, Dacey JWH, Wilhelm LJ, Giovannoni SJ. (2008). SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature* **452**:741–744.
- Vallino JJ, Hopkinson CS, Hobbie JE. (1996). Modeling bacterial utilization of dissolved organic matter: Optimization replaces Monod growth kinetics. *Limnol Oceanogr* **41**:1591–1609.
- Wolf S, Schmidt S, Müller-Hannemann M, Neumann S. (2010). In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinformatics* **11**:148–160.
- Yoch DC. (2002). Dimethylsulfoniopropionate: Its sources, role in the marine food web, and biological degradation to dimethylsulfide. *Appl Environ Microbiol* **68**:5804–5815.
- Zan J, Cicirelli EM, Mohamed NM, Sibhatu H, Kroll S, Choi O, *et al.* (2012). A complex LuxR-LuxI type quorum sensing network in a roseobacterial marine sponge symbiont activates flagellar motility and inhibits biofilm formation. *Mol Microbiol* **85**:916–933.
- Zan J, Liu Y, Fuqua C, Hill RT. (2014). Acyl-homoserine lactone quorum sensing in the Roseobacter clade. *Int J Mol Sci* **15**:654–669.
- Zimmer DP, Soupene E, Lee HL, Wendisch VF, Khodursky AB, Peter BJ, *et al.* (2000). Nitrogen regulatory protein C-controlled genes of *Escherichia coli*: scavenging as a defense against nitrogen limitation. *Proc Natl Acad Sci U S A* **97**:14674–14679.

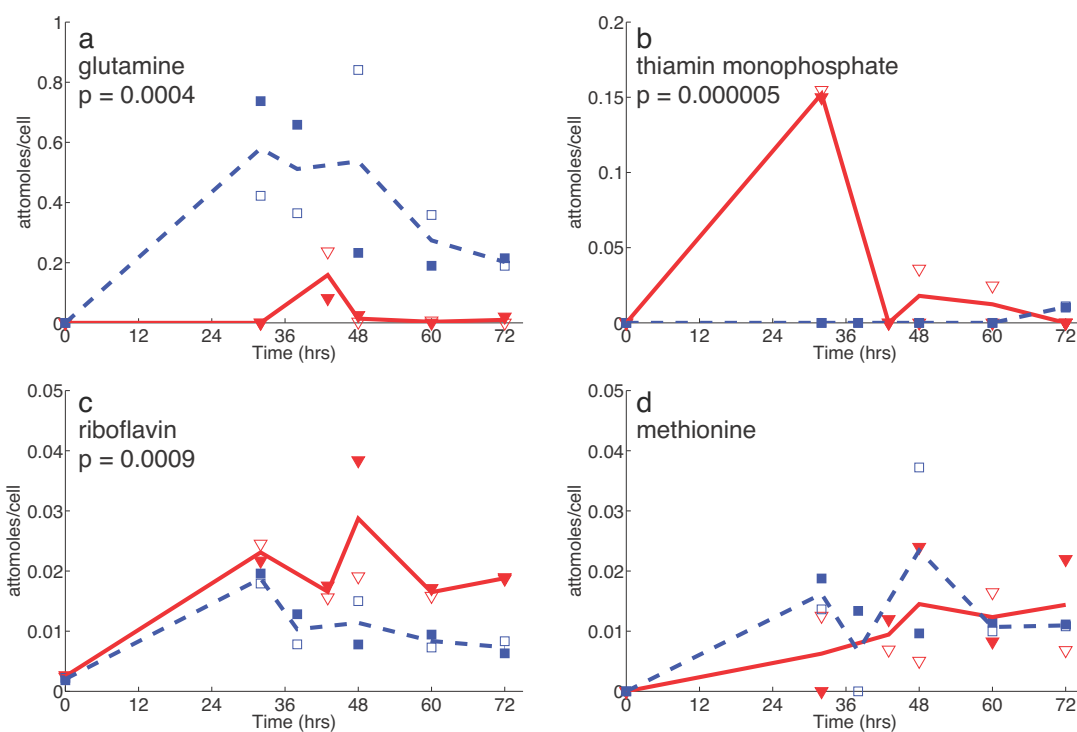


Figure 1: Intracellular metabolite profiles indicate differential concentrations of glutamine and two B vitamins while methionine concentrations are the same. All metabolites presented here were measured in the targeted method: **a**, glutamine. **b**, thiamin monophosphate. **c**, riboflavin. **d**, methionine. Each line reflects the average of two biological replicates, shown in open and closed symbols: red triangles - DMSP treatment; blue squares - propionate treatment. The p -value from an ANOVA test is reported for metabolites with a significant difference in concentration between the substrate treatments across all time points.

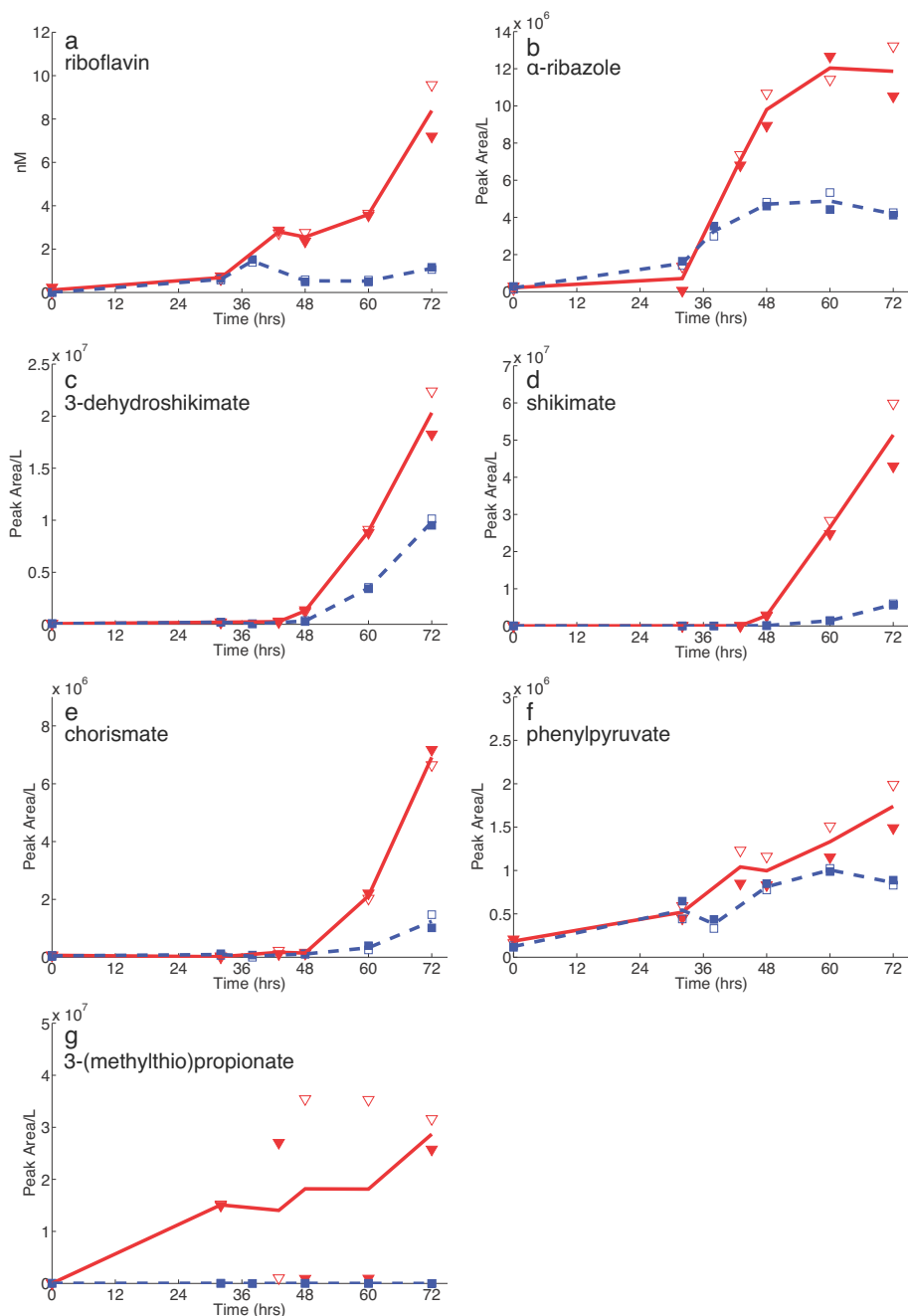


Figure 2: Energetically high-value metabolites are found in the medium at higher concentrations in the DMSP treatment, along with a metabolite associated with DMSP degradation. Each subplot shows the extracellular concentration profile of a metabolite over time. In each subplot, the line reflects the average of two biological replicates, shown in open and closed symbols: red triangles - DMSP treatment; blue squares - propionate treatment. **a**, Riboflavin, measured in the targeted method and reported in nM. The remaining metabolites were measured in the untargeted method and the concentration is reported as peak area / L. **b**, α -ribazole **c**, 3-dehydro shikimate **d**, shikimate **e**, chorismate* **f**, phenylpyruvate* **g**, 3-(methylthio)propionate*. An * denotes putative metabolite identifications with an identification ranking of 3 (see Table S3).

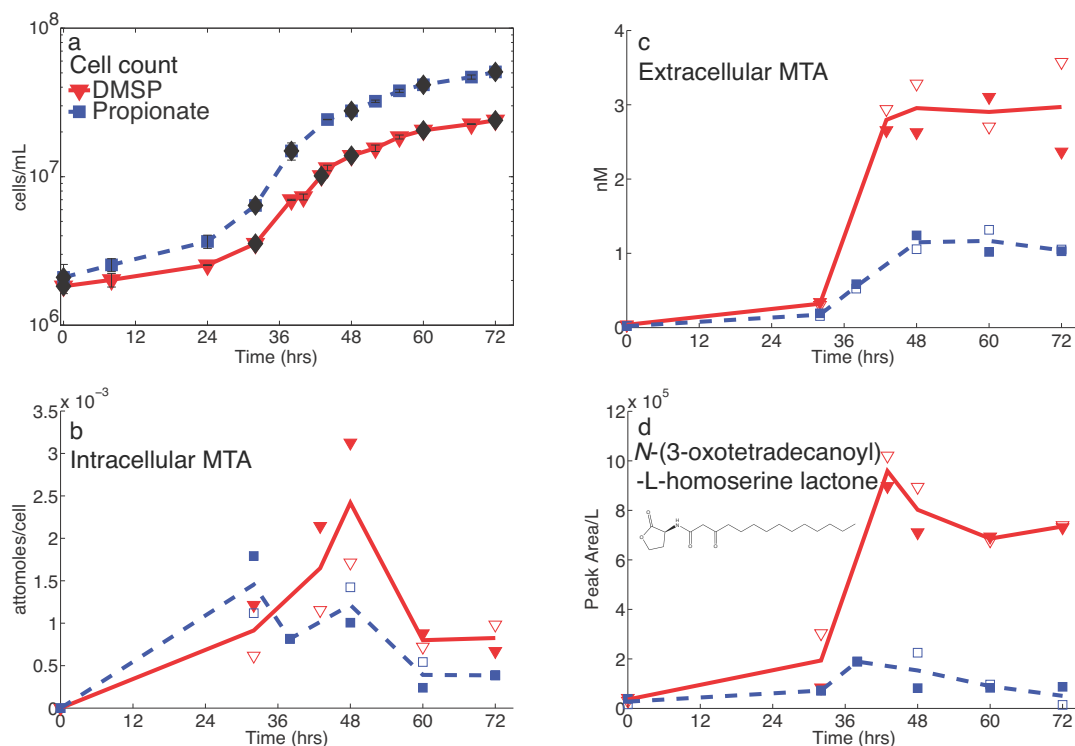


Figure 3: MTA linked to AHL production. **a**, Growth curve for *R. pomeroyi*. Red, solid line is the DMSP treatment and the blue, dashed line is the propionate treatment. Error bars represent one standard deviation of replicate bottles. In the remaining subplots, the line reflects the average of two biological replicates, shown in open and closed symbols: red triangles - DMSP treatment; blue squares - propionate treatment. **b**, The intracellular concentrations of MTA over time, normalized to the number of cells. **c**, The extracellular MTA concentrations over time. **d**, Extracellular concentrations of the quorum sensing compound *N*-(3-oxotetradecanoyl)-L-homoserine lactone. **d-inset**, The structure of *N*-(3-oxotetradecanoyl)-L-homoserine lactone.

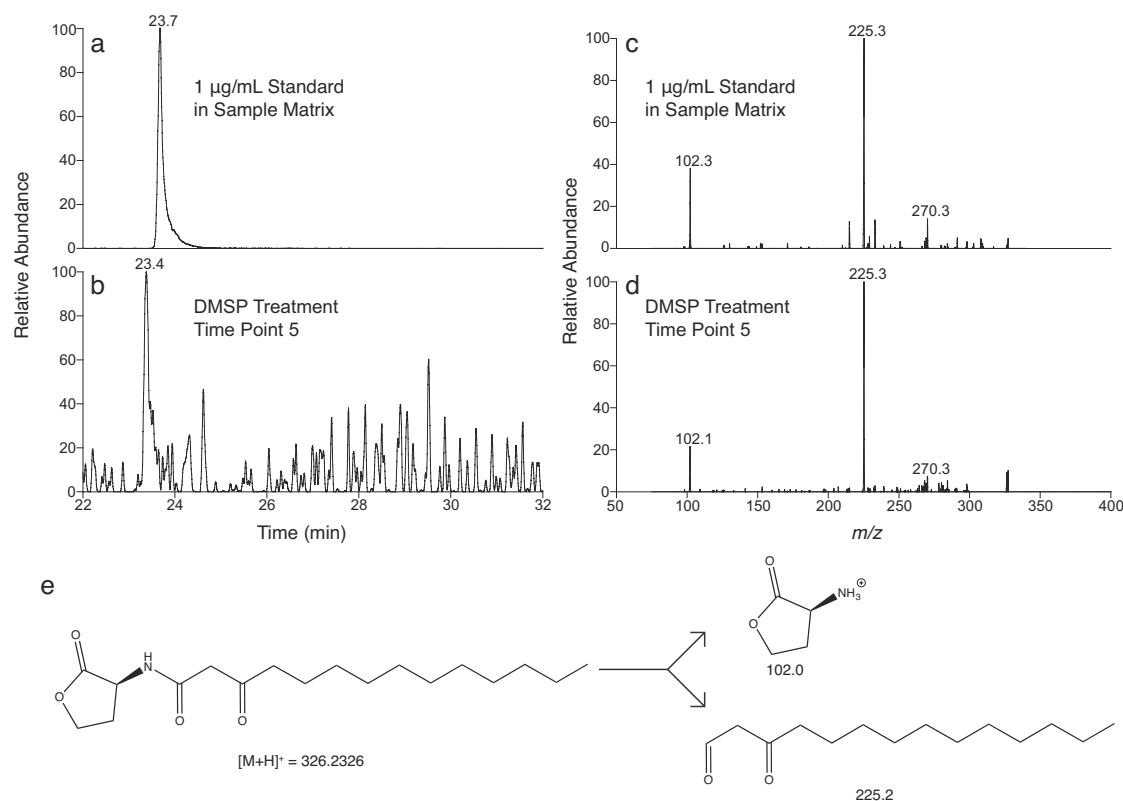


Figure 4: AHL identification. *N*-(3-Oxotetradecanoyl)-L-homoserine lactone commercial standard was spiked into original extracellular extract from the experiment approximately 6 months after the experiment. **a**, Extracted ion chromatograms displaying the mass range within 1 Dalton of the mass of *N*-(3-oxotetradecanoyl)-L-homoserine lactone. **b**, The same data in the original experimental sample. The difference in retention time was confirmed to be due to changes in the column over the 6-month lag between analyses. *N*-(3-Oxotetradecanoyl)-L-homoserine lactone fragments into a lactone head group (102.0 *m/z*) and a tail group fragment (225.2 *m/z*). **c** is the fragmentation of the commercial standard, while **d** is the fragmentation of the feature identified in the experiment. **e**, The structure of *N*-(3-oxotetradecanoyl)-L-homoserine lactone and its primary fragments.

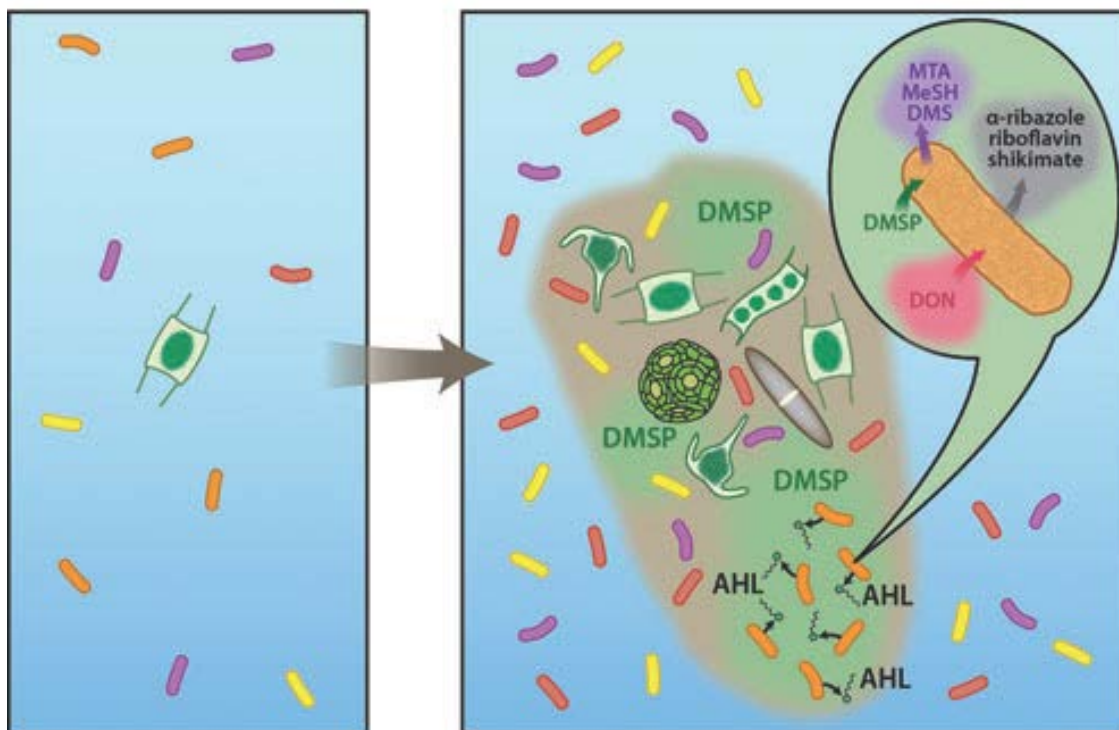


Figure 5: Visual interpretation of *R. pomeroyi*'s response to DMSP. The lefthand panel portrays the diffuse oligotrophic ocean populated by free-living organisms. In contrast, the righthand panel shows dying plankton cells coalescing to form a colloid or particle composed of, and exuding, organic matter. The phytoplankton cells leak DMSP which triggers production of a signaling molecule in *R. pomeroyi* (orange cells). Once the AHL is present at a sufficient concentration, an array of metabolic shifts occur, including reduced intracellular glutamine which may cause increased uptake of dissolved organic nitrogen (DON), increased output of high-value metabolites such as alpha-ribazole, riboflavin, and shikimate, and release of sulfur metabolites such as methanethiol (MeSH; Fig. S1c), dimethylsulfide (DMS; Fig. S1c), and MTA. [image credit: Jack Cook, WHOI]

Supplemental Information

SI Methods

General details. All chemicals were purchased from Sigma-Aldrich unless otherwise indicated. DMSP was purchased from Research Plus, Inc. Methanethiol, ammonium chloride, sodium phosphate, ferrous chloride, ammonium hydroxide, hydrochloric acid (trace metal grade), acetonitrile (optima grade), and methanol (optima grade) were purchased from Fisher Scientific. D3-glutamic acid was purchased from Cambridge Isotopes, d4-4-hydroxybenzoic acid from CDN Isotopes, and d5-sodium taurocholate from Toronto Research Chemicals through Fisher Scientific. We used water purified by a Milli-Q system (Millipore) for all solutions (resistivity 18.2 M Ω •cm @ 25 °C, TOC < 1 μ M). All glassware was acid washed and combusted in an oven at 460°C for at least five hours. All plasticware and media stock solutions were autoclaved before use. In the DMSP and propionate experiment the organic substrates were sterilized by filtering the stock solution through a 0.2 μ m nylon syringe filter. *Ruegeria pomeroyi* DSS-3 [DSM 15171] was purchased from ATCC (Catalog #700808).

Bacterial Cultures. The defined medium was made with 0.2 μ m-filtered seawater collected near high tide from Vineyard Sound (MA) in September 2013. The media was made with either 1 mM sodium propionate or 0.6 mM DMSP (3 mM carbon in both cases), 4 mM ammonium chloride, 30 nM monosodium phosphate, 100 nM ferrous chloride-ethylenediaminetetraacetic acid, 100 nM zinc chloride, 100 nM manganese(II) chloride, 1 nM cobalt(II) chloride, and 1 mL / L medium of f/2 vitamin solution. The bacteria were grown in 250 mL Erlenmeyer flasks (120 mL media) with glass wool in the neck in the dark at 23°C. The experimental samples were inoculated with exponentially-growing *R. pomeroyi* that had been pre-conditioned in the target

media for two transfers. At each time point two duplicate flasks were sacrificed for both treatments. Media blanks for each substrate were extracted at 0, 40, and 72 hours. Cultures were sacrificed at 0, 32, 38(propionate)/43(DMSP), 48, 60, and 72 hours. The third time point was offset between the propionate and DMSP treatments so that the same cell abundance was sampled in each treatment. These time points captured exponential and stationary growth phases in both experiments. Each culture was filtered through a 0.2 μm filter (Omnipore, EMD Millipore) using a glass vacuum filtration rig.

The cell concentration was monitored at 6-10 hour intervals throughout the experiment. Optical density was measured using a Thermo Electron Corporation Nicolet Evolution 300 Spectrophotometer set to 660 nm. This was calibrated to cells/mL using DAPI stained cell counts from cultures grown on the same media to make a five-point calibration curve.

Intracellular Metabolite Extraction (adapted from Rabinowitz and Kimball, 2007). Filters were placed in cryovials and frozen at -80°C immediately after filtration. Subsequently, three quarters of each filter was weighed and cut into small pieces with methanol rinsed scissors on combusted aluminum foil. The pieces of the filter were placed in an 8 mL combusted amber glass vial and 1 mL of -20°C 40:40:20 acetonitrile:methanol:water + 0.1 M formic acid was added to each vial. 25 μL of 1 $\mu\text{g/mL}$ deuterated standard mix (d3-glutamic acid, d4-4-hydroxybenzoic acid, and d5-taurocholate) was added as extraction recovery standards. The vials were vortexed to shake apart filter pieces and fully expose them to the ice-cold solvent. They were sonicated for 10 minutes and the extract was transferred to micro-centrifuge tubes using Pasteur pipettes. The filter pieces were rinsed with 200 μL of extraction solvent and the rinse was added to the micro-centrifuge tubes as well. The extracts were centrifuged at $20,000 \times g$ for

5 minutes. The supernatant was transferred to new 8 mL amber glass vials leaving behind any scraps of filter or cellular detritus. The extracts were neutralized with 26 μ L of 6 M ammonium hydroxide and dried down in a vacufuge. The samples were reconstituted in 495 μ L 95:5 water:acetonitrile and 5 μ L of 5 μ g/mL deuterated biotin (injection standard). The solution (100 μ L) was placed in a glass insert in an autosampler vial for targeted metabolomic analysis.

From this same solution 250 μ L was removed and placed in a new vial, dried down, and brought up to 500 μ L with 0.01 M HCl in water. The solution was pulled through a 50 mg Agilent Bond Elut PPL cartridge that was pre-rinsed with 1 mL methanol. The cartridge was rinsed with 2 mL 0.01 M aqueous HCl, dried for five minutes by pulling air through the cartridge, and then the sample was eluted from the cartridge with 1 mL methanol. The samples were dried down in a vacufuge and reconstituted in 250 μ L of 95:5 water:acetonitrile for untargeted analysis.

The percentage of the filter-extracted carbon retained once it had been re-extracted with a PPL cartridge was 0.1-1.9%. These values reflect the bulk carbon extracted and not the extraction efficiency of individual molecules, which may have higher extraction efficiencies depending on their molecular characteristics.

Extracellular Metabolite Extraction. The media filtrate was acidified to pH 2-3 with 12 M HCl (Dittmar *et al.*, 2008; after modification by Longnecker, 2015). It was then pulled through a methanol-rinsed 50 mg Agilent Bond Elut PPL (styrene divinyl benzene polymer) cartridge. The cartridge was rinsed with 2 mL 0.01 M HCl, and the metabolites were eluted with 1 mL methanol (adapted from Dittmar *et al.*, 2008). It was dried down in a vacufuge and reconstituted to a total volume of 1 mL as the targeted intracellular samples were above. The percentage of

carbon extracted from the media using solid phase extraction was 2.6-18.8%. The low percentages may be related to the fact that most of the dissolved organic carbon in the media was either propionate or DMSP, neither of which are retained well on PPL cartridges (W. Johnson, unpublished results).

Mass Spectrometry. All samples were pre-separated with the same liquid chromatography method on a Phenomenex C18 reversed phase column (Synergi Fusion, 2.1 x 150 mm, 4 μ m). Eluents were (A) Milli-Q water with 0.1% formic acid and (B) acetonitrile with 0.1% formic acid. The gradient included: hold at 5% B for 2 minutes; ramp to 65% B for 16 minutes; ramp to 100% B for 7 minutes and hold for 8 min. The column was re-equilibrated with the starting ratio of eluents for 8.5 min in between analyses.

Untargeted analyses were conducted using a hybrid linear ion trap - 7T Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR-MS; LTQ FT Ultra, Thermo Scientific) with an electrospray ionization source. The instrument was externally calibrated weekly with a standard mix from the manufacturer. In parallel to the high resolution full MS scan in the FT-ICR cell ($R = 100,000$ defined at 400 m/z), the top four most abundant ions that exceeded 500 counts in the ion trap were fragmented. For the intracellular analysis, between every six samples the following sequence was run: Milli-Q blank, a quality control standard mix, Milli-Q blank, two pooled samples (one for conditioning and one for feature validation).

The targeted method used a Thermo Scientific triple quadrupole mass spectrometer with a heated electrospray ionization source. A mixture containing the standards for all of the targeted metabolites was run at 0.5, 1, 5, 10, 50, 100, 250, and 500 ng/mL concentrations to provide an external calibration curve for relative quantification. For the extracellular samples a 500 ng/mL

standard mix was run approximately every six samples as quality control with a Milli-Q blank run before and after. In the intracellular samples a pooled sample of all the experimental samples was run after the second Milli-Q sample to re-condition the column.

Data Processing. The targeted data was processed using Xcalibur (Thermo Scientific software). Standard curves were assessed individually and low quality peaks removed. The media controls were subtracted from the intracellular samples and used as a starting point reference for the extracellular samples.

All untargeted analysis files were converted from Thermo RAW files to mzML files using MSConvert (Chambers *et al.*, 2012). The resulting data was processed using XCMS (Smith *et al.*, 2006; Tautenhahn *et al.*, 2008; Benton *et al.*, 2010) and CAMERA (Kuhl *et al.*, 2012). Peak picking and alignment were carried out separately for the intracellular and extracellular samples. In negative ion mode the peak threshold was set to an absolute intensity of 1000 in MSConvert before processing in XCMS. XCMS settings included: method = centWave, ppm = 2 (negative ion mode), 3 (positive ion mode), signal-to-noise = 10. The retention time correction setting and group.density were also used. Fragmentation mass spectra had a signal-to-noise threshold of 1. For the intracellular samples, pooled samples were used as an additional quality control. The average of each feature in the pooled samples or the experimental samples had to be greater than 1000. The coefficient of variance ($100 \times \text{standard deviation of feature} / \text{mean of feature}$) for each feature in the experimental samples and the pooled samples was calculated and only features that had a higher coefficient of variance in the experimental samples than in the pooled samples were retained for consideration (Vinaixa *et al.*, 2012).

The list of features defined by a retention time and m/z was further refined by removing

any features found in the samples whose average was less than five times higher than the average of all the media blanks. In addition, features had to be present in both biological replicates at two different sampling points. This accounted for compounds that might be found in the seawater but were also being produced by the bacteria. To compare overall differences in metabolite composition in the samples, non-metric multidimensional scaling based on Bray-Curtis distance matrices was used.

Features in the untargeted data, consisting of a unique m/z and retention time, were putatively identified by matching the m/z to metabolites in the METLIN database (Smith *et al.*, 2005) or using MetFrag (Wolf *et al.*, 2010) to match the m/z to metabolites listed in KEGG (Kanehisa *et al.*, 2014). If a fragmentation spectrum was available for a feature, this spectrum was compared to experimentally-generated fragmentation patterns in the METLIN database or to *in silico* derived fragmentation patterns from MetFrag.

DOC & TOC Measurements. 10 mL of whole media was diluted with Milli-Q water to 40 mL and acidified to pH 2-3 for total organic carbon (TOC) analysis. 10 mL of 0.2- μ m filtered media was diluted to 40 mL with MilliQ water and acidified to pH 2-3 for dissolved organic carbon (DOC) analysis. Samples were stored at 4°C until analysis. The non-purgeable organic carbon was measured using a Shimadzu TOC-V_{CSH} Total Organic Carbon Analyzer. A five-point calibration curve made with potassium hydrogen phthalate was used and blanks were run regularly. Comparisons to standards from D. Hansell (University of Miami) were made daily.

Gas Chromatography-PFPD. 2 mL of the unfiltered media was sealed in a serum vial. A 200- μ L sample of headspace was taken with a syringe and injected into a gas chromatograph with an

Alltech AT-Sulfur capillary 0.32 mm ID column and an Ol Corp. pulsed flame photometric detector (adapted from Levine *et al.*, 2012). The presence of methanethiol and dimethylsulfide was confirmed with commercial standards.

Table S1. Total number of features based on m/z and retention time in the untargeted data. Blank-corrected refers to the number of features remaining once the controls were accounted for as described in the supplementary methods. The final column lists the number of blank corrected features that had fragmentation patterns associated with them.

| Ion Mode | | Total | Blank Corrected | Has a Fragmentation Pattern |
|-----------------|---------------|--------------|------------------------|------------------------------------|
| Positive | Intracellular | 9220 | 1117 | 307 |
| | Extracellular | 11724 | 2379 | 448 |
| Negative | Intracellular | 4646 | 897 | 229 |
| | Extracellular | 13543 | 1686 | 396 |

Table S2: Characteristics of intra- and extracellular metabolites observed in two substrate treatments. Columns 2, 3, 4: the number of features found in both substrates compared to the features found in one treatment. Columns 5, 6, 7: the percentage of features at 60 hours in the DMSP treatment whose concentration (in peak area / cell) was 5 times higher than in the propionate treatment, features 5 times higher in the propionate treatment, feature peak area was approximately the same in both treatments (equal), Columns 8, 9, 10: the same data as columns 5-7, at 72 hours.

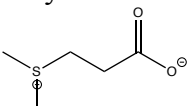
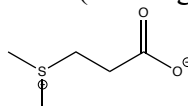
| Ion Mode | Presence/absence | | | Concentrations (Peak area / cell) | | | | | |
|---------------|------------------|---------|------------|-----------------------------------|------------------|-------|------------|------------------|-------|
| Intracellular | | | | | | | | | |
| | | | | 60 hours | | | 72 hours | | |
| | DMSP | Overlap | Propionate | DMSP 5x | Propionate 5x | Equal | DMSP 5x | Propionate 5x | Equal |
| Neg | 4% | 43% | 53% | 9% | 86% | 5% | 6% | 86% | 8% |
| Pos | 1% | 96% | 3% | 17% | 64% | 19% | 13% | 66% | 21% |
| Extracellular | | | | | | | | | |
| | | | | 60 hours | | | 72 hours | | |
| | DMSP | Overlap | Propionate | DMSP 5x | Propionate 5x | Equal | DMSP 5x | Propionate 5x | Equal |
| Neg | 1% | 98% | 1% | 19% | 28% | 53% | 24% | 43% | 33% |
| Pos | 6% | 89% | 5% | 23% | 36% | 41% | 22% | 51% | 27% |

Table S3. Matches of m/z and fragments for putatively identified compounds discussed in the paper. Column 2: the exact mass adjusted for either negative ion mode ($[M-H]^-$) or positive ion mode ($[M+H]^+$), Column 3: measured m/z of the feature in the untargeted data, Column 4: the ppm difference between the feature m/z and the adjusted exact mass, Column 5: the measured retention time (RT) of the feature in minutes, Column 6: possible fragments calculated from MetFrag that matched those of the feature, Column 7: fragments and their intensity relative to the highest peak (after the comma) from the METLIN database that were generated in an Q-TOF mass spectrometer, Column 8: Measured fragments of the feature observed in the experiment and their intensity relative to the highest peak (after the comma), Column 9: The confidence ranking of identification based on the reporting standards proposed by Sumner et al. (2007) where 1 is the highest confidence identification.

| Compound | Exact mass | Feature m/z | ppm match | RT | Predicted MetFrag fragments | METLIN | Measured fragments | Confidence Ranking |
|-----------------------------------------------------|------------|---------------|-----------|---------------------------|-----------------------------|--------------------------------------------------------------|-------------------------------------------------------------|--------------------|
| 3-dehydro shikimate | 171.0299 | 171.0300 | 0.6 | 4.8 | 127.0 83.0 141.0 | 171.0, 100 127.0, 26 109, 7 81.0, 1 | 127.0, 100 83.0, 21 102.9, 9 141.0, 3 171.0, 50 | 2 |
| shikimate | 173.0455 | 173.0457 | 1.2 | 7.7 | 129.0 127.0 111.0 | 173.0, 100 93.0, 9 111.0, 7 137.0, 4 129.0, 0 | 129.0, 100 127.0, 39 111.0, 15 173.0, 70 | 2 |
| chorismate | 225.0405 | 225.0406 | 0.4 | 8.3 | -- | -- | -- | 3 |
| phenylpyruvate | 163.0401 | 163.0402 | 0.6 | 10.2 | -- | -- | -- | 3 |
| 3-(methylthio) propionate | 119.0172 | 119.0173 | 0.8 | 7.8 int/ 7.2 ext | -- | -- | -- | 3 |
| <i>N</i> -(3-oxotetradecanoyl)-L-homoserine lactone | 326.2326 | 326.2327 | 0.3 | 23.7 | -- | 326.2, 100 102.1, 97 225.2, 76 | 327.0, 7 101.9, 24 225.1, 100 270.1, 9 | 1 |
| α -ribazole* | 279.1339 | 279.1340 | 0.4 | 6.6 | 147.0 | -- | 147.0, 100 | 2 |

*Fragmentation pattern from literature also supports this identification (O'Toole *et al.*, 1994).

Table S4. Initial intermediates produced by DMSP degradation. There are two pathways for DMSP degradation: demethylation/demethiolation and cleavage (Reisch *et al.*, 2011, 2013). *R. pomeroyi* has been demonstrated to utilize both pathways (González *et al.*, 2003). The demethylation/demethiolation pathway produces methanethiol which can be lost as a gas or incorporated into biomass. The cleavage pathway cleaves dimethylsulfide first before utilizing the remainder of the molecule. Both pathways ultimately lead to central carbon metabolism. The table summarizes the intermediates detected in each pathway. “Detected?” column indicates whether the metabolite was detected in any of our methods.

| DMSP (Demethylation/Demethiolation) | | DMSP (Cleavage) | |
|-----------------------------------------------------------------------------------|--------------|-------------------------------------------------------------------------------------|----------------|
|  | |  | |
| Intermediate | Detected? | Intermediate | Detected? |
| 3-(methylthio)propionate | Yes (no MS2) | acrylate | No (too small) |
| 5-methyl-THF | No | Dimethylsulfide (DMS) | Yes |
| MMPA-CoA | No | acryloyl-CoA | No |
| MTA-CoA | No | | |
| acetaldehyde | No * | | |
| Methanethiol (MeSH) | Yes | | |

*I Acetaldehyde will not ionize under our ESI conditions and it is too small to be detected by the FT-ICR MS with our parameters.

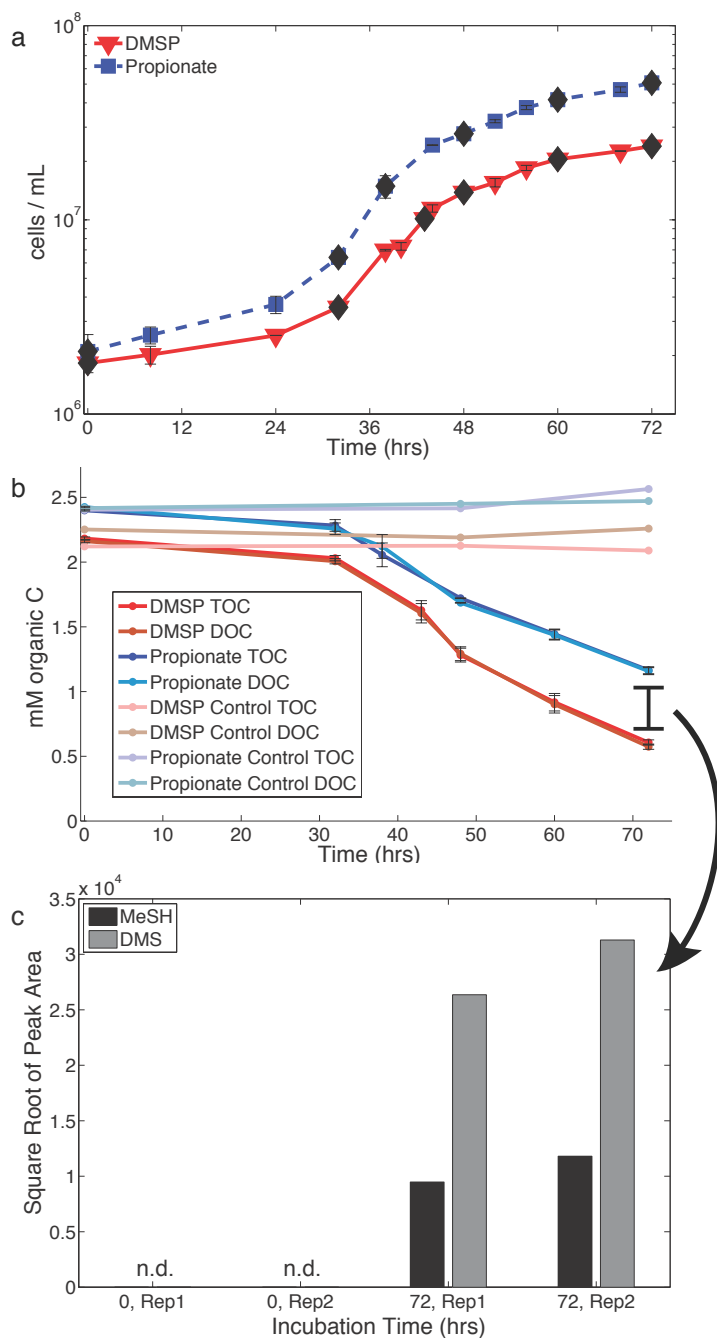


Figure S1. *R. pomeroyi* growth, carbon remineralization, and organosulfur gas production.

a, Average growth curves of *R. pomeroyi* growing on DMSP (red) and propionate (blue). The black diamonds indicate the time points where sampling occurred. The error bars are the standard deviation for the two biological replicates. It is unclear why there is an offset in cell density between the two treatments but is perhaps due to different metabolic efficiencies for catabolism of the two carbon substrates. **b**, Total organic carbon (TOC; dark red, DMSP treatment and dark blue, propionate treatment) and dissolved organic carbon (DOC; orange, DMSP treatment and light blue, propionate treatment) concentrations. Cell-free media controls are in faded colors (red-orange, DMSP and blue propionate) and show no decline in organic

carbon during the experiment. Error bars represent one standard deviation for the biological replicates. The target concentration for both carbon substrates was 3 mM but some loss of the substrates may have occurred with filter sterilization. **c**, Abundance of dimethylsulfide (DMS) and methanethiol (MeSH) in the headspace of the initial DMSP cultures and after 72 hours. The presence of both gases confirms that both the demethylation pathway and the lyase pathway (see Table S4) were used by *R. pomeroyi* for DMSP catabolism. In **S1b** the DMSP treatment has less organic carbon at the outset than propionate and that difference increases by the final timepoint at 72 hours. Subtracting the initial difference still leaves a difference in both TOC and DOC of 0.3 mM at 72 hours. This difference is due to the loss of DMS and MeSH as shown in **S1c** that only occurs in the DMSP treatment. If we calculate the total loss of DOC over the course of the experiment for each substrate, the DMSP treatment lost 1.4 mM carbon and the propionate treatment lost 1.1 mM carbon. However, to compare bacterial utilization, it may be more relevant to consider the moles of each molecule respired rather than the total carbon because the portion of DMSP that enters primary carbon metabolism is the same 3-carbon unit as propionate. DMS and MeSH are lost from the medium as gases. We can approximate this by normalizing to the number of carbon atoms in each molecule (5 for DMSP and 3 for propionate). We calculate a loss of 0.3 mM for DMSP and 0.4 mM for propionate. Thus the draw-down of moles of each organic substrate is not very different, especially assuming that there would be some difference in efficiency between the pathways.

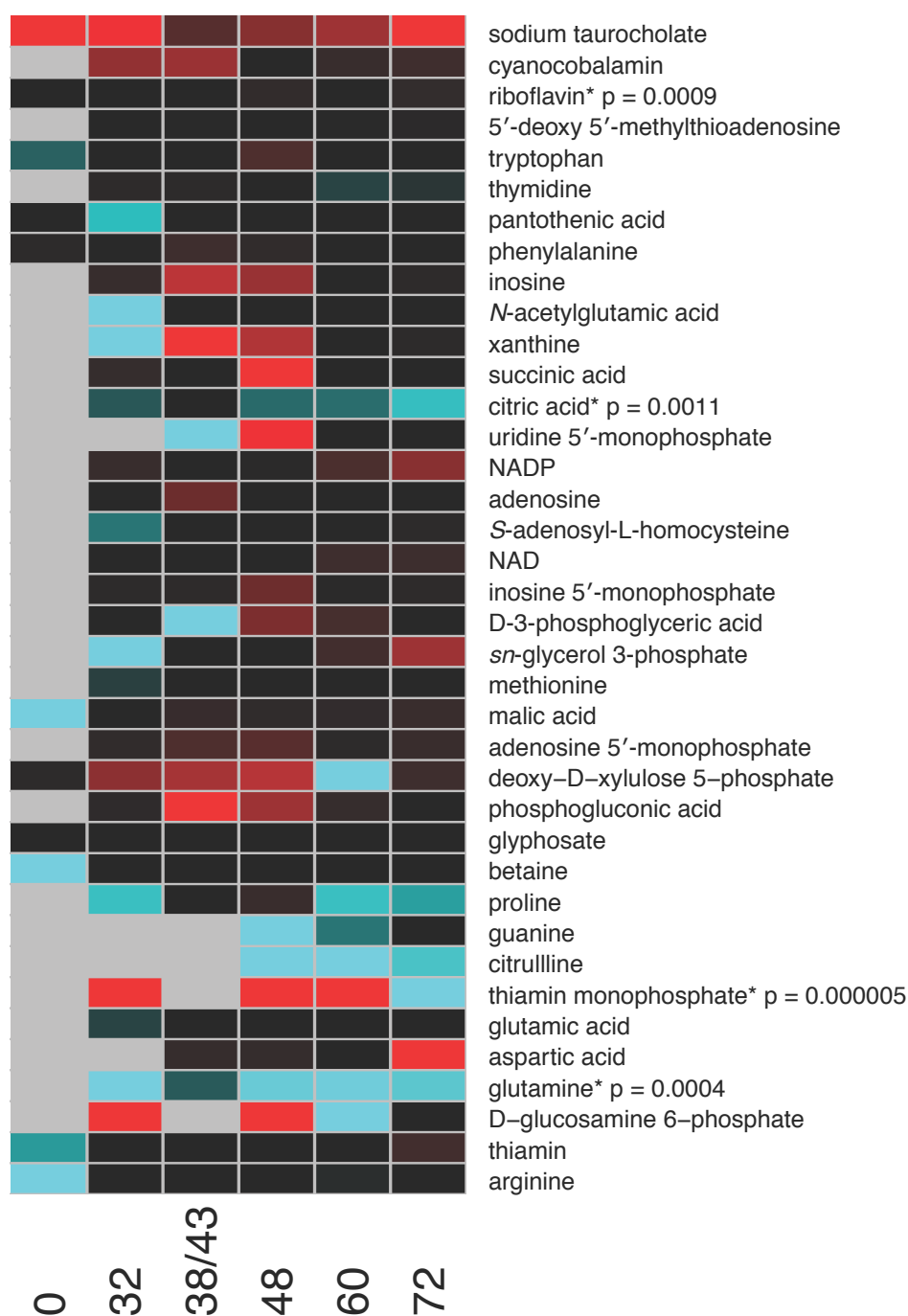


Figure S2. A heatmap of all the targeted intracellular metabolite concentrations in the two treatments. Averages of biological replicates are compared at each time point (0 hours-72 hours). (Red): concentrations are higher in the DMSP treatment increasing from dark to light; (Blue) concentrations were higher in the propionate treatment increasing from dark/greenish to cyan; (Black) the concentrations in the two substrate treatments were similar; (Gray) metabolites were not detected in either treatment. Metabolites with a significant difference in concentration between the substrate treatments are indicated with (*) and the *p*-value from an ANOVA test is reported.

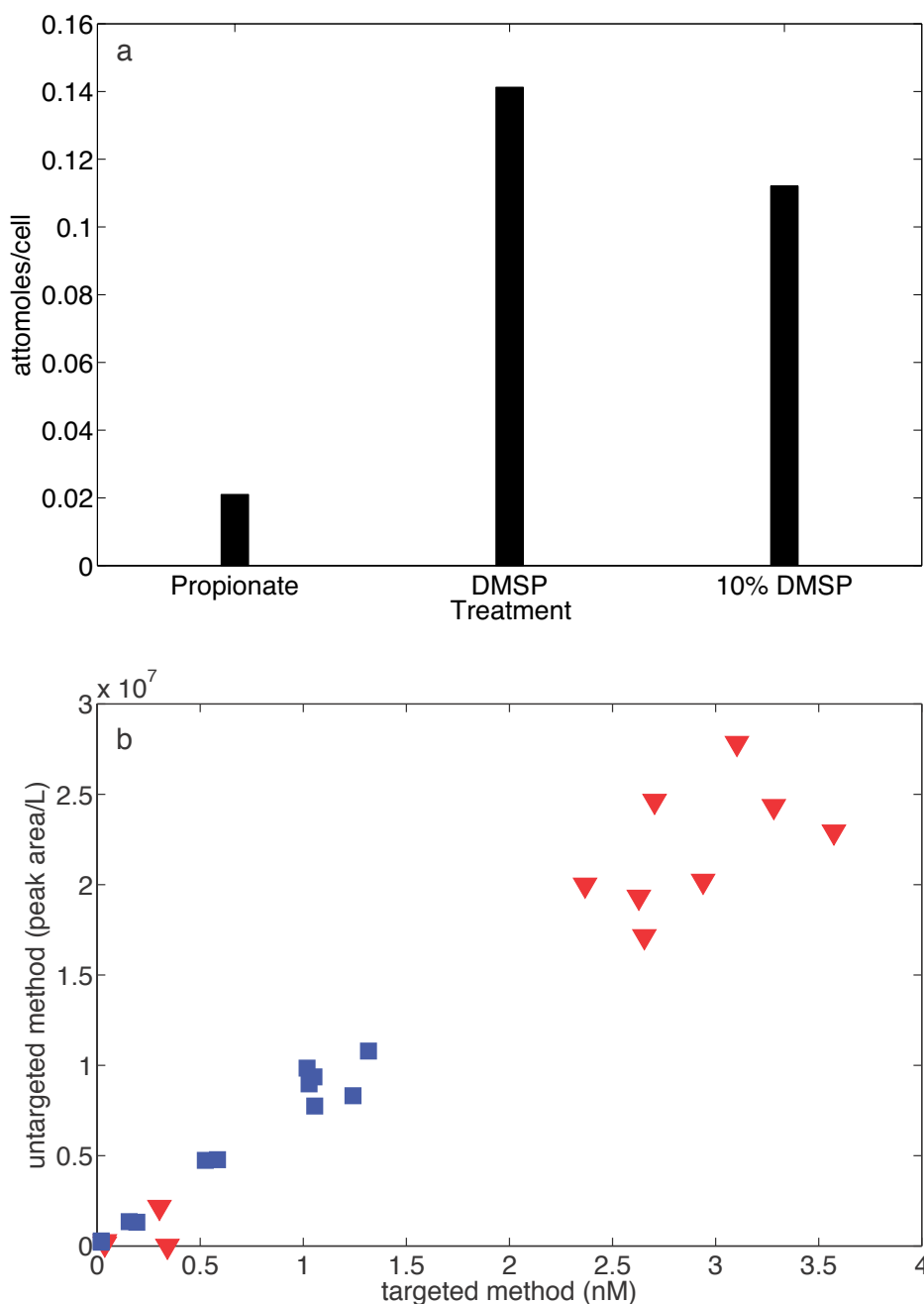


Figure S3. MTA concentrations in the 90% propionate / 10% DMSP experiment and the relationship between MTA response in the targeted and untargeted methods. a, Extracellular concentrations of MTA during stationary growth phase, normalized to cell abundance. *R. pomeroyi* was grown on media with 100% propionate, 100% DMSP, and 10% DMSP / 90% propionate. The total carbon added to each treatment was the same (3 mM). **b,** Comparison of extracellular MTA measured in the targeted method versus the untargeted method. The linear trend indicates that relative MTA response was similar on both instruments. There was a significant correlation between the two methods (Pearson correlation coefficient, $r = 0.9778$, $p < 0.0001$). Red is the DMSP treatment, blue is the propionate treatment.

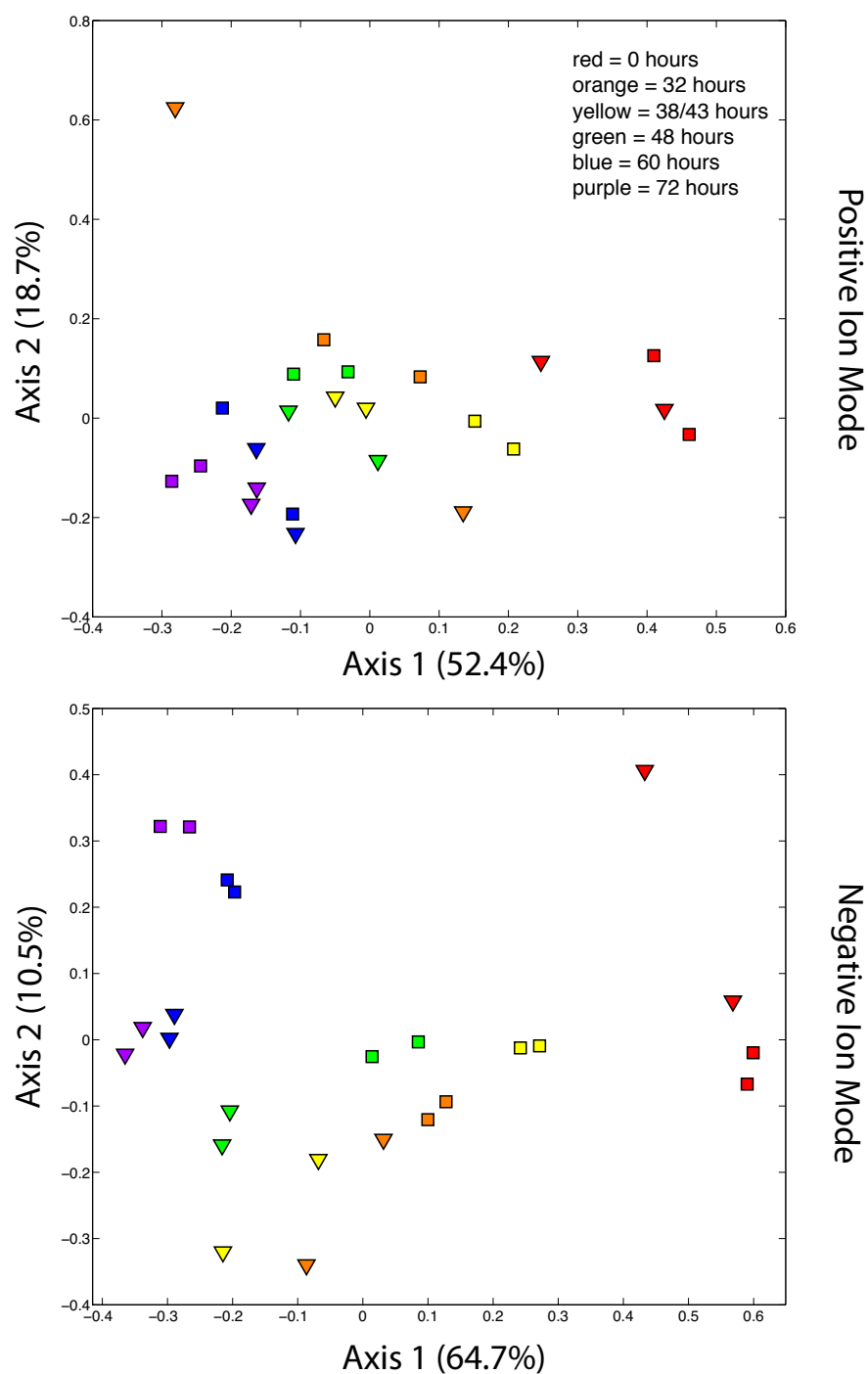


Figure S4. Non-metric multidimensional scaling plots for extracellular samples based on a Bray-Curtis distance matrix of peak area normalized to filtrate volume. For both positive and negative ion modes, the dominant factor is time although there is some clustering based on substrate. The relative percent of variation explained by each axis is included in the label. Each marker represents a sample and more similar samples appear closer to one another in this plot. Squares represent propionate samples and triangles represent DMSP samples.

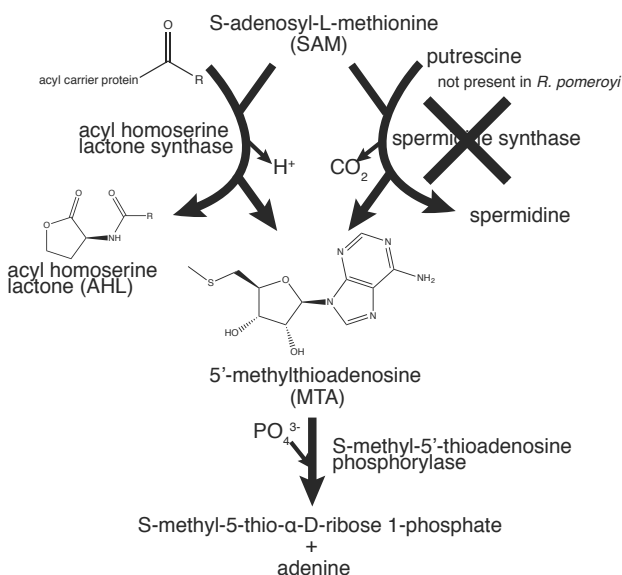


Figure S5. Overview of MTA production and degradation. MTA is a by-product of two metabolic pathways. One produces an acyl homoserine lactone (left) and the other converts putrescine to spermidine (right). MTA and an AHL were measured extracellularly at higher concentrations in the DMSP treatment. The *R. pomeroyi* genome lacks this pathway for spermidine/spermine synthesis that produces MTA and instead contains the genes for an alternate pathway requiring the enzymes carboxyspermidine dehydrogenase and carboxyspermidine decarboxylase to convert putrescine to spermidine. Only the first gene (S-methyl-5'-thioadenosine phosphorylase) of the classic methionine salvage pathway has been annotated in the *R. pomeroyi* genome.

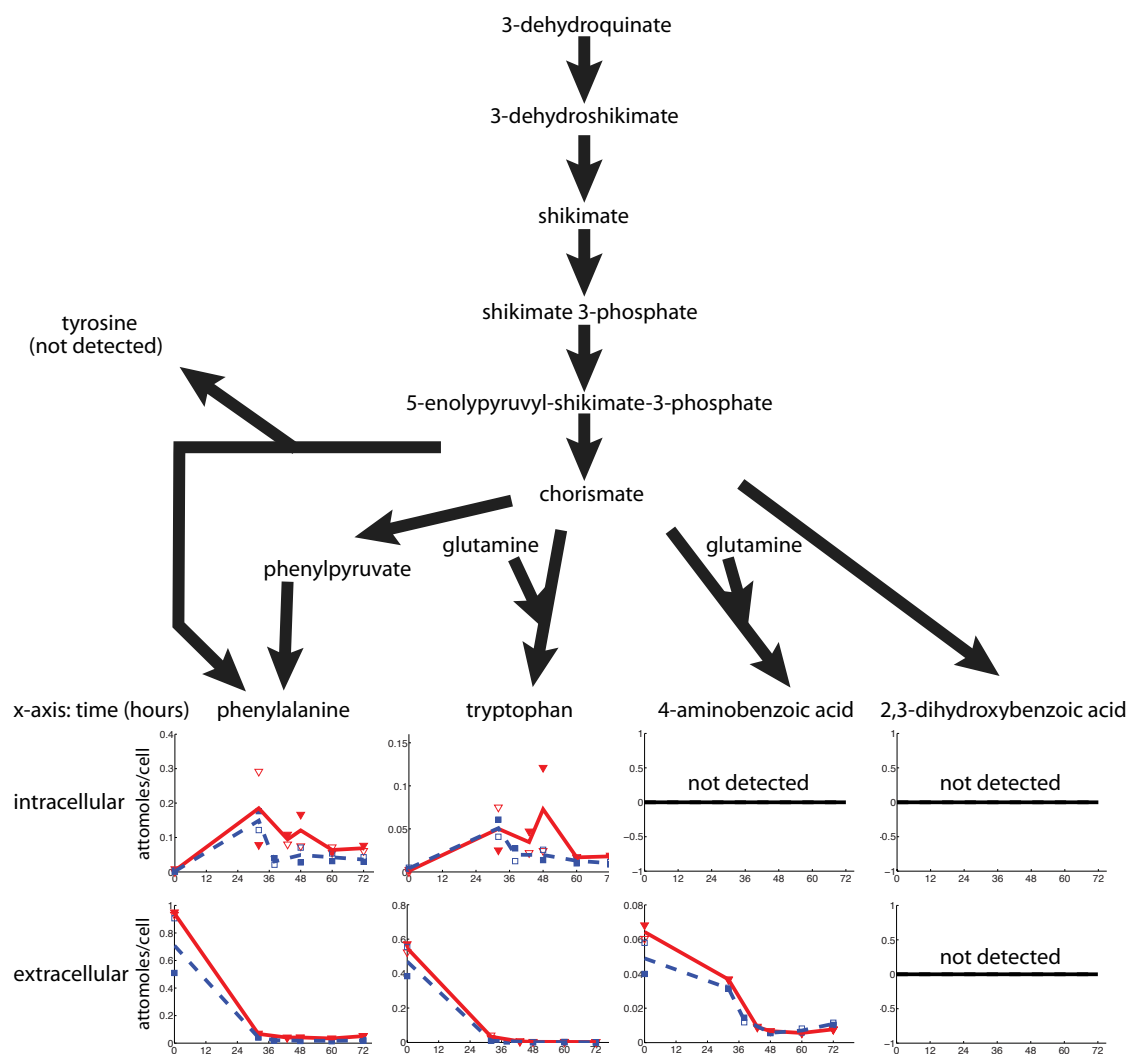


Figure S6. End-products of the biosynthetic pathways that require chorismate and chorismate biosynthesis. Tyrosine was not included in the targeted method and was not detected in the untargeted method. Phenylalanine, tryptophan, 4-aminobenzoic acid, and 2,3-dihydroxybenzoic acid were all quantified in the targeted method. If present, the cell-normalized concentrations were similar between both treatments. Red: DMSP treatment, Blue: propionate treatment, Black: compound was not detected in either treatment.

Supplemental References

- Benton HP, Want EJ, Ebbels TMD. (2010). Correction of mass calibration gaps in liquid chromatography-mass spectrometry metabolomics data. *Bioinformatics* **26**:2488–2489.
- Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, *et al.* (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **30**:918–920.
- Dittmar T, Koch B, Hertkorn N, Kattner G. (2008). A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol Oceanogr Methods* **6**:230–235.
- González JM, Covert JS, Whitman WB, Henriksen JR, Mayer F, Scharf B, *et al.* (2003). *Silicibacter pomeroyi* sp. nov. and *Roseovarius nubinihibens* sp. nov., dimethylsulfoniopropionate-demethylating bacteria from marine environments. *Int J Syst Evol Microbiol* **53**:1261–1269.
- Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* **42**:199–205.
- Kuhl C, Tautenhahn R, Bottcher C, Larson TR, Neumann S. (2012). CAMERA: An integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal Chem* **84**:283–289.
- Levine NM, Varaljay VA, Toole DA, Dacey JWH, Doney SC, Moran MA. (2012). Environmental, biochemical and genetic drivers of DMSP degradation and DMS production in the Sargasso Sea. *Environ Microbiol* **14**:1210–1223.
- Longnecker K. (2015). Dissolved organic matter in newly formed sea ice and surface seawater. *Geochim Cosmochim Acta* **171**:39–49.
- O’Toole GA, Trzebiatowski JR, Escalante-Semerena JC. (1994). The cobC gene of *Salmonella typhimurium* Codes for a novel phosphatase involved in the assembly of the nucleotide loop of cobalamin. *J Biol Chem* **269**:26503–26511.
- Rabinowitz JD, Kimball E. (2007). Acidic acetonitrile for cellular metabolome extraction from *Escherichia coli*. *Anal Chem* **79**:6167–6173.
- Reisch CR, Crabb WM, Gifford SM, Teng Q, Stoudemayer MJ, Moran MA, *et al.* (2013). Metabolism of dimethylsulphoniopropionate by *Ruegeria pomeroyi* DSS-3. *Mol Microbiol* **89**:774–791.
- Reisch CR, Moran MA, Whitman WB. (2011). Bacterial catabolism of dimethylsulfoniopropionate (DMSP). *Front Microbiol* **2**:172–184.
- Smith CA, Maille GO, Want EJ, Qin C, Trauger SA, Brandon TR, *et al.* (2005). A metabolite mass spectral database. *Ther Drug Monit* **27**:747–751.
- Smith CA, Want EJ, O’Maille G, Abagyan R, Siuzdak G. (2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **78**:779–787.
- Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, *et al.* (2007). Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group

(CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3**:211–221.

Tautenhahn R, Böttcher C, Neumann S. (2008). Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* **9**:504–520.

Vinaixa M, Samino S, Saez I, Duran J, Guinovart JJ, Yanes O. (2012). A guideline to univariate statistical analysis for LC/MS-based untargeted metabolomics-derived data. *Metabolites* **2**:775–795.

Wolf S, Schmidt S, Müller-Hannemann M, Neumann S. (2010). In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinformatics* **11**:148–160.

Chapter 4.

Insights into the controls on metabolite distributions along a latitudinal transect of the western Atlantic Ocean

Abstract

Metabolites, or the small organic molecules that are synthesized by cells during metabolism, are essential currency in the vast array of interactions that occur within a microbial community. Characterizing the distribution of these molecules both inside marine microbial cells and dissolved in seawater is essential to understanding the controls on their production and fate, as well as their roles in structuring microbial communities. Here, we apply a targeted metabolomics method to characterize particulate and dissolved distributions of a suite of biologically relevant small molecules along a latitudinal transect in the western Atlantic Ocean. These molecules include sugars involved in primary carbon metabolism, vitamins, amino acids, nucleic acids, osmolytes, and intermediates in biosynthetic pathways. While many metabolite distributions co-vary with biomass abundance, other metabolites show distributions that suggest abiotic, species specific, or metabolic controls on their variability. This study applies targeted metabolomics to a large transect in the ocean, and this comprehensive overview of metabolite distributions highlights metabolites that may have particular significance in microbial adaptations and interactions.

4.1. Introduction

Marine organic matter is a complex component of the global carbon cycle due to its molecular diversity. Cycling of this organic matter is largely controlled by the relationship between microbial metabolism and dissolved organic carbon (DOC) as marine microbes produce, consume, and influence the composition of DOC. About 50% of the organic matter fixed by phytoplankton passes through the marine DOC pool (Ducklow 1999), which is a carbon reservoir similar in size to that of carbon dioxide in the atmosphere. Within the marine DOC pool, organic molecules are cycled on timescales ranging from minutes to thousands of years (reviewed by Carlson and Hansell 2015). To adequately understand the factors that control production and removal of organic matter, molecular-level characterization of this material is needed. Compositional studies of organic matter have predominantly focused on compound classes, determining that a large portion of high-molecular weight dissolved organic matter (DOM) is made up of carbohydrates (Aluwihare et al. 1997) as well as carboxyl-rich aliphatic material (Hertkorn et al. 2006). Hydrolysable and free individual amino acids make up a smaller portion of DOC at only around 1-4% of DOC (McCarthy et al. 1996; Kaiser and Benner 2009). An additional group of molecules in DOM, uncovered with ultrahigh-resolution mass spectrometry of solid phase extracted DOM, includes compounds with molecular formulas that do not match common biochemical compound classes (Dittmar and Koch 2006). Small, highly labile biomolecules, while a relatively small proportion of DOM in the ocean, play essential roles as the currency in marine microbial interactions, transferring energy, cellular components, and information among microbial neighbors. By simultaneously measuring a suite of these biological molecules, we can trace the relationships between their distributions in particulate and dissolved phases and hypothesize about the controls on the production and removal of these molecules and

their roles in microbial communities.

These small organic biomolecules, or metabolites, are defined as the products of all cellular regulation (Fiehn 2002). Due to the sensitivity of metabolite concentrations to both genetic and environmental factors, the metabolite profile of an organism can be considered an aspect of its phenotype. Intracellular metabolite concentrations respond to physical, chemical, and biological environmental cues including nutrient limitation (Brauer et al. 2006), salinity (Gebser and Pohnert 2013), temperature (Thompson et al. 1992), oxidative stress (Lesser 2006), grazing (Pohnert 2000), viral lysis (Ankrah et al. 2014), and the presence of infochemicals (Seyedsayamdost et al. 2011). For example, in response to infection by a phage, a marine bacterium had elevated intracellular concentrations of some amino acids and sugars (Ankrah et al. 2014). Understanding how metabolite concentrations respond to these environmental cues and interactions allows us to assess, not only how microbes adapt physiologically to the environment but also, how environmental stressors can impact the microbial organic matter that is released into the dissolved phase.

Dissolved metabolites mediate interactions among marine organisms. This encompasses both the transfer of organic substrates through the foodweb as well as the subtler but often significant role of infochemicals in influencing an organism's phenotype. Sugars are a dominant component of the marine foodweb, fueling aerobic respiration (Søndergaard et al. 2000). However, heterotrophic microbes also catabolize other types of metabolites. For example, dimethylsulfoniopropionate (DMSP), an algal metabolite, is used as both a source of energy and reduced sulfur (Kiene et al. 2000). Amino acids also move through the foodweb, providing both fixed nitrogen and essential cellular building blocks (Rich et al. 1997). These interactions are not constrained to the autotrophic production of organic substrates for heterotrophic

remineralization. For instance, certain species of phytoplankton require vitamins that they cannot synthesize *de novo*, such as vitamin B₁₂ or biotin (Croft et al. 2006). Some heterotrophic bacteria also have specific substrate requirements, such as a SAR11 isolate that requires an exogenous source of a precursor to thiamin (Carini et al. 2014). These physiological demands have been studied on a case-by-case basis using field and lab experiments coupled with genomic information. However, our knowledge of the distributions of many of these metabolites in the ocean is limited or non-existent, thus inhibiting our ability to predict the impact these requirements may have on microbial communities.

To quantify a suite of structurally distinct metabolites from a single sample, we use a liquid chromatography-tandem mass spectrometry-based metabolomics approach. In this targeted metabolomics method, pure standards of each metabolite of interest are used to identify and quantify that metabolite in the sample. Samples were collected from latitudes ranging from 55°N to 38°S in the western Atlantic and at depths from 5 m to ~ 5500 m, encompassing a variety of ocean regions. The method applied to these samples allowed us to measure 59 particle-associated metabolites and 26 dissolved metabolites. These metabolites include amino acids, vitamins, nucleic acids, osmolytes, and a variety of metabolic intermediates from both primary metabolism and biosynthetic pathways. Many metabolites are essential for most organisms, while others have been identified in the course of laboratory experiments as being responsive to environmental differences or being specific to certain groups of organisms. Some of these metabolites have been measured previously in the ocean, allowing us to compare our measurements to literature values while other metabolites have never been measured, to our knowledge, in the ocean.

The goal of this chapter is to quantify the distribution of a core set of metabolites along

latitudinal and depth gradients in the Atlantic Ocean. In particular, identification of metabolites whose distributions deviate from that of microbial biomass can indicate distinct controls on the production or degradation of those molecules in certain ocean regions. The majority of the metabolites currently included in the targeted metabolomics method are essential components of metabolism that are required for any living organism to function. Thus, starting from the null hypothesis that each metabolite is present in the same abundance inside a cell anywhere in the ocean and that its release into, and uptake from, the dissolved pool of metabolites is driven only by the number of cells present, particulate and dissolved metabolite distributions should co-vary with biomass. However, we know that this is not the case, and that metabolites vary in response to a wide variety of factors. To identify metabolites that have differential responses in the ocean, metabolite distributions will be discussed in the following contexts: 1) metabolite distributions that are dissociated from biomass in the euphotic zone, 2) particulate metabolite composition in the euphotic zone versus the deep ocean, and 3) the relationship between particulate and dissolved distributions of metabolites and the implications for their rates of cycling.

4.2 Materials and methods

4.2.1 Field Sites

Samples were collected over the course of two cruises in 2013. Cruise KN210-04 (Deep DOM) took place in the western tropical Atlantic from 25 March – 9 May (austral fall), transiting from Montevideo, Uruguay to Bridgetown, Barbados. In the North Atlantic, samples were collected on the second leg of cruise AE1319 transiting from Boothbay Harbor, Maine U.S.A. to Bermuda via the Labrador Sea from 20 August – 11 September (boreal late summer – early fall). On KN210-04 samples were collected at depths including 5 m (referred to as the

surface), the deep chlorophyll maximum (DCM) determined by fluorescence, 250 m, Antarctic Intermediate Water (AAIW, ~1000 m), North Atlantic Deep Water (NADW, ~2500 m), and Antarctic Bottom Water (AABW, ~5000 m). On cruise AE1319 samples were collected at 5 m, the DCM, Eighteen Degree Mode Water (~350 m, where present), ~1000 m, and 3000 m.

4.2.2 Materials

All metabolite standards were obtained from Sigma-Aldrich at the highest purity available with the exceptions of dimethylsulfonylpropionate (DMSP), which was purchased from Research Plus, Inc., 2,3-dihydroxypropane-1-sulfonate (DHPS) and acetyltaurine, which were donated by Dr. Mary Ann Moran (University of Georgia), and *S*-(1,2-dicarboxyethyl)glutathione, which was purchased from Bachem. Hydrochloric acid (trace metal grade), acetonitrile (Optima grade), and methanol (Optima grade), were purchased from Fisher Scientific. Formic acid was obtained from Fluka Analytical. Glutamic acid- d_3 was purchased from Cambridge Isotopes, 4-hydroxybenzoic acid- d_4 from CDN Isotopes, and sodium taurocholate- d_5 from Toronto Research Chemicals through Fisher Scientific. Water was purified by a Milli-Q system (Millipore; resistivity $18.2 \text{ M}\Omega\cdot\text{cm}$ @ 25°C , TOC $< 1 \text{ }\mu\text{M}$). Glassware used to collect samples and GF/F filters were combusted in an oven at 460°C for at least five h before being brought on the cruises and all plasticware was autoclaved before use. The filter holders and tubing for shipboard filtration were flushed with 10% HCl and then Milli-Q water between each sample.

4.2.3 Shipboard sample processing

Water (4 L) was collected directly from Niskin bottles into polytetrafluoroethylene (PTFE) or polycarbonate bottles. Water was subsequently filtered through a $0.7 \text{ }\mu\text{m}$ GF/F filter (Whatman) and a $0.2 \text{ }\mu\text{m}$ filter (Omnipore, EMD Millipore) using a peristaltic pump. GF/F filters

were re-wrapped in their combusted aluminum foil envelopes and the omnipore filters were folded and placed into cryogenic vials (Nalgene). Filters were stored at -80°C until they could be extracted in the laboratory. The filtrate was then acidified with 4 mL of 12 M HCl (~pH 2-3; Dittmar et al. 2008; Longnecker et al. 2015). Dissolved organic molecules were extracted from the filtrate using solid phase extraction (SPE) modified styrene-divinylbenzene polymer (Agilent Bond Elut PPL) (see Chapter 2 for extraction efficiencies). The PPL cartridges were first rinsed with 6 mL of methanol and then the filtrate was pulled through the cartridge via PTFE tubing using a vacuum pump (Dittmar et al. 2008). The cartridge was rinsed with ~24 mL of 0.01 M HCl and then allowed to dry by pulling air over the cartridge for 5 min. The sample was then eluted with 6 mL of methanol into a glass test tube and transferred with a Pasteur pipette to an 8 mL amber vial. Extracts were stored at -20°C until analysis.

4.2.4 Laboratory sample processing

Filters were extracted within 48 h prior to mass spectrometry analysis. Filter extraction was adapted from Rabinowitz and Kimball (2007). One half of each filter was weighed and cut into small pieces using methanol rinsed scissors and tweezers on combusted aluminum foil. The pieces were placed in an 8 mL glass amber vial with 1 mL of -20°C extraction solvent (40:40:20 acetonitrile:methanol:water + 0.1 M formic acid). 25 µL of a 1 µg/mL deuterated standard mix (glutamic acid-d₃, 4-hydroxybenzoic acid-d₄, taurocholate-d₅) were spiked in as an extraction recovery standard. Each vial was gently vortexed to separate filter pieces and expose them to the solvent. The vials containing the filter pieces and solvent were then sonicated for 10 min. The solvent extract was transferred with a Pasteur pipette to 1.5 mL Eppendorf centrifuge tubes. The filter pieces left in the 8 mL vials were rinsed with 200 µL of cold extraction solvent and that was added to the centrifuge tubes. The extract was spun at 20,000 x g for 5 min to remove

cellular detritus and filter particles. Supernatant was transferred to clean 8 mL glass amber vials. Extracts were neutralized with 26 μL of 6 M ammonium hydroxide. Solvent was removed by vacufuge until samples were almost completely dry ($<5 \mu\text{L}$). They were then reconstituted with either 247.5 μL 95:5 water:acetonitrile and 2.5 μL biotin- d_2 (injection standard) for samples collected at the surface or DCM or 123.75 μL 95:5 water:acetonitrile and 1.25 μL biotin- d_2 (injection standard) for deeper samples with presumed lower biomass. 100 μL of each solution were loaded into a glass insert in an autosampler vial. 15 μL of each sample were combined to create a pooled sample.

For the dissolved metabolites, 1 mL of the SPE extracts ($\sim 6 \text{ mL}$ total volume, but the exact volume of each sample was measured with a glass syringe) were brought to almost complete dryness in the vacufuge and then reconstituted in 495 μL 95:5 water:acetonitrile and 5 μL of 5 $\mu\text{g/mL}$ biotin- d_2 (injection standard). Some dissolved metabolite samples (mostly from deep samples, 2000 – 5000 m) required additional dilution due to ion suppression in the middle of the chromatogram (see Table S1 for a list of these samples and what dilution was used).

4.2.5 Mass spectrometry

The LC-MS/MS analysis was performed on a Phenomenex C18 column (Synergi Fusion, 2.1 x 150 mm, 4 μm) coupled via heated electrospray ionization (ESI) to a triple quadrupole mass spectrometer (Thermo Scientific TSQ Vantage) operated under selected reaction monitoring mode (SRM), as described previously (Kido Soule et al. 2015). Quantification and confirmation SRM transitions were monitored for each analyte. Eluent A (Milli-Q water with 0.1% formic acid) and Eluent B (acetonitrile with 0.1% formic acid) were run at 250 $\mu\text{L} / \text{min}$ in the following gradient: hold at 5% B for 2 min; ramp to 65% B for 16 min; ramp to 100% B for 7 min and hold for 8 min. The column was re-equilibrated with the starting ratio of eluents for 8.5

min between sample analyses. Samples were run in batches of approximately 50 with their pooled samples comprised of those 50 samples. Before each batch of samples was run, the column was conditioned with 5 injections of the pooled samples. A pooled QC sample was then run after every ten samples.

4.2.6 Data processing

XCalibur RAW files generated by the mass spectrometer were converted to mzML files using MSConvert (Chambers et al. 2012). MAVEN (Melamud et al. 2010; Clasquin et al. 2012) was used to select and integrate peaks. Calibration peaks below a MAVEN quality threshold of 0.4 (on a scale of 0-1) and sample peaks below 0.2 were discarded. To enhance confidence in metabolite identification, quantification and confirmation peaks were required to have retention times within 12 seconds (0.2 minutes) of each other. Peak widths can range from 10-30 sec and thus, with the lower intensity of the confirmation ion peak, its assigned retention time can deviate from that of the quantification ion. Confirmation ions were required to have a MAVEN quality score of at least 0.1 and a signal-to-noise ratio greater than 1. Calibration curves were required to have at least five calibration points, with the highest point at one concentration level above the highest concentration in a sample. Metabolite abundances were normalized to the volume. Where noted, particulate concentrations of metabolites were normalized to the total moles of targeted metabolites measured in the sample because this parameter correlated well with available cell counts (see discussion below and Figure S2). The concentrations of dissolved metabolites with extraction efficiencies greater than 1% on a PPL cartridge were corrected to reflect a more accurate estimate of the *in situ* metabolite concentrations (see extraction efficiencies in Chapter 2 (Johnson et al., submitted)). Dissolved metabolites with extraction efficiencies less than 1% were not reported.

4.2.7 Total organic carbon (TOC)

40 mL samples of seawater were collected in combusted glass EPA vials. They were acidified to pH 2-3 with concentrated hydrochloric acid and stored at 4°C until analysis. They were analyzed on a Shimadzu TOC-VCSH total organic carbon analyzer coupled to a TNM-1 analyzer. Milli-Q water blanks and standard curves of potassium hydrogen phthalate were run throughout analysis and comparisons were made daily to standards from Prof. D. Hansell (University of Miami).

4.2.8 Temperature, salinity, chlorophyll *a*, and cell counts

The R/V *Knorr* was equipped with a SBE9+ CTD with a depth limit of 6000 m. A SBE3T/SBE4C sensor system was used to measure temperature and conductivity. A Wet Labs FLNTURTD combination fluorometer and turbidity sensor was used to detect fluorescence. The R/V *Atlantic Explorer* has a SBE 9/11 Plus CTD with a depth limit of 6800 m with a SBE 3plus temperature sensor and a SBE 4C conductivity sensor. A Chelsea Aquatracka II was used to measure fluorescence. Fluorescence data from both cruises were calibrated to direct chlorophyll *a* measurements. Chlorophyll *a* samples from the R/V *Knorr* were measured by Gwenn Hennon according to methods from Arar & Collins (1997). Chlorophyll *a* samples from the R/V *Atlantic Explorer* were provided by Michael Lomas and measured according to methods from Yentsch and Menzel (1963). Erin Eggleston provided prokaryotic cell counts using epifluorescence microscopy of cells stained with SYBR Green I stain and captured on a 0.02 µm anodisc filter (for complete method see Noble and Fuhrman 1998).

4.2.9 Data validation

To check for possible leakage of metabolites from burst cells on filters into the dissolved sample, metabolites that were measured in the dissolved phase were plotted as the particulate

ratio of metabolite x / phenylalanine against the dissolved ratio of metabolite x / phenylalanine (Figure S1). Phenylalanine was chosen because it was measured in most dissolved and particulate samples, thus providing a ubiquitous denominator for this comparison. If metabolites were leaking, the ratio in the particulate should equal the ratio in the dissolved for any given sample, as presumably there is no selectivity in which metabolites leak from burst cells. All the metabolites examined have many samples that were measured only in the particulate or only the dissolved fraction, thus reducing the number of comparisons that can be made, and supporting the contention that leakage was not likely to be a significant problem. Where this comparison could be made, in some cases, riboflavin, MTA, and inosine in particular, had samples that seem to have similar particulate and dissolved ratios. However, this was highly variable across the different metabolites suggesting that there was no systemic release of these metabolites into the dissolved phase, as would be expected from cell leakage. We conclude that cell leakage, or breakage during filtration, had a minimal impact on our data.

4.2.10 Computation and statistical tools

Matlab (R2014a; MathWorks, Natick, MA) was used to process data and create some figures. The map and metabolite profile images were created with Ocean Data View (Schlitzer 2016). Spearman's rank correlation coefficients were calculated using the R stats package (R Core Team 2015) to examine the relationship between metabolite distributions and also metadata. Spearman's rank correlation identifies monotonic relationships but does not require that the relationship be linear.

4.3 Results

4.3.1 Oceanographic transect

Samples were collected along a transect encompassing latitudes from 55°N to 38°S and at depths ranging from the surface (5 m) to approximately 5500 m. Surface oceanic regions along this transect included the South Atlantic Gyre (Stns K2, K7, K9, K11), the equatorial region (Stns K15, K19), the Amazon plume (Stn K23), the North Atlantic Gyre (Stns A11, A15), and the Labrador Sea (Stns A4, A7; Figure 1a) facilitating an analysis of metabolite distributions in both productive and oligotrophic regions of the ocean. In addition, samples were taken from deep water masses including Antarctic Intermediate Water, North Atlantic Deep Water, and Antarctic Bottom Water. Stations with particularly high productivity can be observed in the chlorophyll *a* concentrations (Figure 1c). These included the deep chlorophyll maximum (DCM) near the equator where there was upwelling of nutrients (Stn K15) and the highest latitudes sampled in the Labrador Sea (Stns A4, A7) where there was a shallow DCM. The total organic carbon (TOC) concentrations, as expected, were maximal in the surface water of the subtropical gyres and decrease with depth (Figure 1b).

4.3.2 Data normalization

Both particulate and dissolved metabolite concentrations are reported per total volume filtered in picomolar units. This approach is the most unambiguous way to present the data as concentration units are the standard chemical and geochemical approach to reporting chemical data. Concentration units also facilitate comparisons between the particulate and dissolved distributions. However, as particulate metabolite samples are approximations of the intracellular pool of metabolites, an approach that presents metabolite abundances in relation to their abundance inside of cells is also needed. Other ‘omics fields use approaches such as normalization to housekeeping genes (whose abundances are constant inside cells) to present variability on a per cell basis. We have too little data on metabolite abundances and variability in

microbes at this point in time to have identified a “housekeeping” metabolite. Other oceanographic parameters were considered as proxies for biomass such as chlorophyll *a* concentration or the beam transmittance at 660 nm (from a transmissometer) as a measure of how many particles were in the water. Chlorophyll *a* is problematic because it is only produced by phytoplankton, and its abundance can be affected by physiological changes (Kruskopf and Flynn 2006). The relationship between transmissometer data and prokaryotic cell counts (from SYBR Green staining and counting under an epifluorescence microscope) showed a poor relationship. However, these cell counts correlate well ($r^2 = 0.9014$; Figure S2) with the total moles of targeted particulate metabolites measured. The total abundance of targeted particulate metabolites is distributed in a way that is consistent with expectations for productive versus oligotrophic oceanic regions (Figure S3). While cell counts only include prokaryotes, we can expect that prokaryotic abundances are roughly linked to phytoplankton abundances (Bird and Kalff 1984). Thus, normalizing to the total targeted metabolites roughly removes the confounding factor of biomass. This normalization is also similar to approaches previously used in oceanography for amino acids, where the mole fraction of the individual amino acids are reported relative to the total amount of amino acids measured (Lee and Cronin 1984; Dittmar et al. 2001). Thus this type of normalization gives a sense of the abundance of a metabolite in this pool of targeted metabolites. As more measurements are made of these metabolites in the ocean, best practices in normalization will inevitably emerge.

4.3.3 Variability in metabolite detection

Of the 59 metabolites that passed the quality control checks in the particulate samples, 12 were not detected in any sample. 2,3-dihydroxybenzoic acid, desthiobiotin, indole 3-acetic acid, inosine 5'-monophosphate, cysteine, glutathione, *S*-(5'-adenosyl)-L-homocysteine, and *sn*-

glycerol 3-phosphate were each measured in only one sample in the entire transect preventing assessment of their variability across the full transect. Most metabolites were elevated in the upper 200 m. Dimethylsulfoniopropionate (DMSP), guanine, and phenylalanine were measured in every particulate sample above 200 m. Guanine and phenylalanine were the only metabolites measured in all but 3 samples throughout the entire water column (Table 1).

Of the metabolites that were not detected in any particulate samples, some, like fosfomycin and glyphosate, are most likely not produced by many microbes and when they are, their concentrations may be quite low. It was surprising, however, to observe no detectable concentrations of metabolites such as cytosine, thymidine, and taurine. Cytosine is a pyrimidine nucleobase, while thymidine is a pyrimidine nucleoside. Thus, both are essential molecules in all living organisms. While purine metabolites, like guanine, were detected at high concentrations in the particulate samples throughout the water column, cytosine and thymidine must have much lower concentrations, < 2 pM and < 10 pM, respectively. The lack of taurine is surprising because it is an osmolyte, and these molecules are generally present at relatively high intracellular concentrations. Despite a low detection limit of 0.4 pM, it was not detected in any samples. For these metabolites, larger sample volumes may be required to measure them in the ocean.

There were 26 dissolved metabolites that passed quality control and could be extracted using a PPL solid phase extraction method (Johnson et al., submitted). Of these, 8 were not detected in any samples. In some cases this may be due to extremely low solid phase extraction efficiencies (i.e., high detection limits). No metabolites were measured in every sample in the dissolved phase. However, phenylalanine was the most prevalent of the dissolved metabolites detected in these samples. 4-Aminobenzoic acid, pantothenic acid, and tryptophan were present

in most samples above 200 m (Table 2).

4.3.4 Trends in particulate metabolites

Many metabolites measured in the particulate fraction were strongly associated with the euphotic zone (Table 1, see number of samples with metabolites at depths < 200 m). This is consistent with the higher productivity and more abundant biomass in the euphotic zone. For example, guanine, a purine nucleobase required by all living organisms, is measured in nearly all the samples in the transect allowing comparison of its concentrations at all depths. Particulate guanine abundance (picomoles / liter, pM) is relatively high in the surface samples, generally peaks at the DCM, and then is quite low at depth (Figure 2a). As an essential component of DNA, this is a metabolite that should roughly correlate with cell abundance, and is observed to do so at all the stations in this transect. When guanine is normalized to the total moles of targeted metabolites measured in a sample, it becomes a major component of these deep samples reaching greater than 80% of the total moles measured in some samples (Figure 2b).

To compare metabolite distributions along the transect, Spearman's rank correlation coefficients were used to determine the extent of positive or negative co-variation in the distributions of metabolites. Particulate metabolites were compared in this way both normalized to volume (Figure 3a) and normalized to total moles (Figure 3b). Comparison of metabolites normalized to volume highlight differences in the distribution of metabolites in the euphotic zone (see Table 1 for the number of samples in which each metabolite is detected in the upper 200 m versus below 200 m). In figure 3a, the particulate metabolites in group B vary approximately with biomass abundance in the euphotic zone while metabolites that do not vary with biomass abundance are found in group A. To support this, an additional correlation plot with total moles included as a proxy for biomass shows these two groups (Figure S4). Adenosine 5'-

monophosphate (AMP) is representative of the distributions seen for group B metabolites (Figure 4a) and shows a strong linear ($r^2 = 0.7099$) correlation with chlorophyll *a* concentrations (Figure 4b). Metabolites in this group (B) are characterized by elevated concentrations at high latitudes in the North Atlantic, at the DCM at K15 where upwelling was observed, and at the surface of K23 where there was some influence from the Amazon River Plume. Relatively low concentrations were measured in the surface samples for stations K7, K9, and K11, which are all in the South Atlantic subtropical gyre.

Comparison of metabolites normalized to total moles (Figure 3b) distinguishes those metabolites that were exclusively measured in the euphotic zone from those that persisted with depth. Metabolites in group 1 were measured in deep samples while metabolites in group 2 were measured predominantly in the euphotic zone. Some metabolites in group 1, including guanine, phenylalanine, tryptophan, and caffeine, are measured throughout the deep ocean while other metabolites are more variable in their distribution (Figure 3b).

4.3.5 Trends in dissolved metabolites

The distributions of dissolved metabolites are linked to biomass in some cases, but others have distinct euphotic zone distributions (Figure 5). The correlation plot focuses only on samples collected above 200 m as there were too few metabolites measured in the deeper samples (Table 2). In the dissolved fraction there are three groups of metabolites that separate with the greatest distance between them in the dendrogram. Here, the focus will mainly be on the Group B and C metabolites. First, group C metabolites have a positive correlation with chlorophyll *a* suggesting they are linked more strongly to biomass, particularly phytoplankton. Group C metabolites include MTA, tryptophan, riboflavin, phenylalanine, 2,3-dihydroxybenzoic acid, and pantothenic acid. For example, riboflavin is higher in many of the DCM samples relative to the surface

samples and the highest overall concentrations are in the Labrador Sea where the highest chlorophyll *a* concentrations are also observed (Figure 6a). In contrast, group B metabolites correlate more strongly with parameters such as TOC and temperature both of which are highest at the surface and lower latitudes suggesting an inverse pattern for these metabolites compared to those associated with chlorophyll *a*. These metabolites include 4-hydroxybenzoic acid, caffeine, (iso)leucine, and 4-aminobenzoic acid. For example, 4-hydroxybenzoic acid is only measured in one DCM sample but is measured in nearly all of the surface samples particularly in the South Atlantic (Figure 6b). Notably, the two northernmost stations in the Labrador Sea with the highest surface chlorophyll *a* concentrations did not contain 4-hydroxybenzoic acid.

4.3.6 Comparison to previously measured concentrations and residence times

Previous work has shown that the recovery of isotopically labeled standards added at the beginning of the filter extraction protocol yields near 100% recovery (Kido Soule et al. 2015). However, to evaluate whether the measured particulate and dissolved metabolite abundances were consistent with each other, an estimated dissolved metabolite residence time was calculated where possible (see Table S2 for values and calculation). In some cases the values were expected, on the order of a 1-2 day residence time but many were higher, reaching 100 days (Table S2). As a comparison, dissolved free amino acids have had their uptake and turnover rates measured using radiotracers. Their turnover rates are generally on the order of hours to several days (reviewed by Nagata 2008). However, because these turnover rates are calculated by measuring total amino acid incorporation into biomass (both free and protein-associated amino acids), the particulate fraction in these studies is a larger pool of amino acids than was measured here. If the residence times estimated here were adjusted to account for the larger pool of free and protein-incorporated amino acids, they would be much shorter. On sinking particles total

hydrolysable particulate amino acids can be about 1000 times greater in abundance than free amino acids (Lee and Cronin 1982) although this may be exaggerated somewhat by preferential degradation of free amino acids. The residence time could also be longer than expected if there is incomplete extraction of the metabolites from the filters. Incomplete recovery, however, is difficult to assess.

Some of the metabolites presented here have been measured previously; in some cases using similar mass spectrometry-based methods as well as with other analytical approaches. While the abundance of particulate metabolites is difficult to compare due to the diversity of ways in which molecules are normalized, concentrations of dissolved metabolites can be compared more readily. Of the 26 metabolites that are possible to characterize using our current method, concentrations in aquatic systems were found for 11 of them (Table S3). Of these, six metabolites were detected in at least one sample, allowing a comparison. Two B vitamins, biotin and riboflavin, have values of the same order of magnitude as the measurements made by Sañudo-Wilhelmy et al. (2012). A third B vitamin, thiamin, has an order of magnitude higher concentration than that reported by Sañudo-Wilhelmy et al. (2012), but is only measured in one sample. Thiamin has a low extraction efficiency in the current protocol (see Chapter 2) and the resulting correction may introduce error. The amino acids, (iso)leucine, phenylalanine, and tryptophan, when measured as dissolved free amino acids, display a variety of concentrations ranging from hundreds to thousands of picomolar (Mopper and Lindroth 1982). In the method used here, concentrations varied from < 10 pM to hundreds of pM. The lower sample concentrations in this dataset come from the deep ocean, whereas the comparison is being made to samples taken from 1-170 m in the Baltic Sea. The surface ocean concentrations in this dataset are at the low-end of the range measured by Mopper and Lindroth in the upper water column in

the Baltic Sea (1982). In short, dissolved metabolite concentrations obtained using this method agree within an order of magnitude with existing data, where available.

4.4 Discussion

The metabolites whose oceanic distributions are presented here are, for the most part, central to cellular function across the domains of life. If marine microbial life functioned in a fairly homogeneous way throughout the ocean, we might expect these molecules to vary simply as a function of biomass, but our results indicate that this is not the case for all metabolites. In the euphotic zone, examples of metabolites in both the particulate and dissolved phases that deviate from this distribution are indicative of metabolic variability across oceanic regions. Differences in metabolite composition between the euphotic zone and the deep ocean can also be instructive as possible indicators of organic matter degradation or distinct metabolisms at depth. Finally, the relationship between particulate and dissolved distributions of individual metabolites indicates varied cycling of these molecules.

4.4.1 Differential metabolite distributions in the euphotic zone

In the euphotic zone many particulate metabolites have similar profiles (Figure 3a, group B) most likely due to having relatively constant concentrations inside of cells. However, the metabolites in group A deviate from that type of distribution (Figure 3a). These include two B vitamins, a number of amino acids, and some nucleobases and their derivatives.

For example, adenine, a purine nucleobase, is measured in most surface samples, but in relatively few DCM samples compared to metabolites in group B. It is particularly striking that it is not detected at the Stn K15 DCM where nutrients were being upwelled, but that some of its highest concentrations are measured in the surface at Stns K5 and K7 which were located in the oligotrophic South Atlantic subtropical gyre. This distinct pattern in particulate adenine

distribution in the ocean may be linked to how microbes respond to nutrient stress. In culture, *Escherichia coli* responds to both carbon and nitrogen starvation by elevating intracellular concentrations of adenine (Brauer et al. 2006). While the mechanism for this increase is not known, intracellular adenine concentrations play a regulatory role inside of cells by binding to certain riboswitches (reviewed by Winkler and Breaker 2005). However, this trend is not entirely consistent, for instance, there was no adenine detected at Stn K9 which is also in the gyre and the highest adenine concentration in the transect is measured in a surface sample from Stn K23 where there was influence from the Amazon River plume. A diatom bloom at this station suggests that this was a relatively nutrient replete station. Thus, particulate (or intracellular) adenine concentrations may be driven by something more complex than simply nutrient limitation, such as type or quality of nutrients, or the response may be species-specific.

Interestingly, arginine and (iso)leucine are also both observed to increase in *E. coli* under carbon and nitrogen stress, but the fold increase is not as great as for adenine (Brauer et al. 2006). However, their distributions differ from that of adenine along this transect. Caffeine and biotin were detected in so few particulate samples that it is difficult to conclude anything about their distribution. Pyridoxine, a B vitamin like biotin, is measured in more samples but does not demonstrate a clear trend. Uracil, a pyrimidine base in RNA, is undetected in a number of samples in the Labrador Sea where chlorophyll *a* concentrations were highest but, unlike adenine, it is absent from the surface samples in the South Atlantic subtropical gyre. Uracil is found to accumulate in *E. coli* cells grown in batch culture when they reach stationary growth phase, and this has been attributed to the possible degradation of rRNA (Rinas et al. 1995), which might occur during cell death or nutrient stress. Finally, aspartic acid and ornithine are both amino acids that are intermediates in the urea cycle, which distributes nitrogen in the cell,

but they do not display similar distributions in this dataset. These distributions that deviate from biomass may highlight subtle and perhaps highly transient differences in metabolic state that are not reflected in more conservatively maintained intracellular metabolites.

Of the dissolved metabolites, 4-aminobenzoic acid, (iso)leucine, 4-hydroxybenzoic acid, and caffeine correlate with TOC concentrations and temperature (Figure 5, group B), while biotin and chitotriose are linked to salinity. (Iso)leucine, biotin, and chitotriose were not measured extensively throughout the transect and show a scattered distribution. Both 4-aminobenzoic acid and 4-hydroxybenzoic acid were particularly elevated in surface samples. 4-aminobenzoic acid was also measured in the DCM but 4-hydroxybenzoic acid was not.

Dissolved caffeine, surprisingly, was measured throughout the euphotic zone. It does not appear to be a contaminant from the ship because it was detected not only at the surface but also at the DCM and as deep as 250 m. The concentrations ranging from 2.6-8.7 pM were on the low end of the range that has been measured in coastal waters (Weigel et al. 2002; Rodriguez del Rey et al. 2012). Interestingly, these studies sought to link caffeine concentrations along the coast of Oregon (Rodriguez del Rey et al. 2012) and in the North Sea (Weigel et al. 2002) to sources of pollution but found that there was no link between caffeine concentrations and pollutant sources, while other pollutants could be clearly linked to human inputs. This surprising, unexplained variability suggests that there could be another source of caffeine not considered in these studies. Specifically, terrestrial plants synthesize caffeine as a defense compound (Mithöfer and Boland 2012); perhaps there are phytoplankton that can do the same. In fact, caffeine has been measured intracellularly in cultures of the diatom, *Thalassiosira pseudonana* (Krista Longnecker, personal communication). Taken together, these data suggests that there may be a microbial source, albeit small, of caffeine in the ocean.

In metabolism, 4-hydroxybenzoic acid is a precursor to quinone and quinol biosynthesis. Quinones and quinols are used in transmembrane electron transport. However, 4-hydroxybenzoic acid is also used to measure the photo-oxidative production of hydroxyl radicals because benzoic acid reacts with the hydroxyl radical to form 4-hydroxybenzoic acid (Vione et al. 2010). This indicates that 4-hydroxybenzoic acid can be generated through photo-oxidation of organic molecules. Additionally, 4-hydroxybenzoic acid can be an oxidation product of lignin degradation. However, unlike other phenols that are derived from lignin oxidation, 4-hydroxybenzoic acid can be generated from non-vascular plants and the non-woody portion of vascular plants. Chemical oxidation experiments have shown that the amino acid, tyrosine, is a likely source of 4-hydroxybenzaldehyde and likely also of the acid and ketone forms (Hedges and Parker 1976). Natural DOM sources have been shown to generate 4-hydroxybenzoic acid as well as smaller amounts of the ortho and meta isomers of the molecule during photo-oxidation (Page et al. 2011). The fact that 4-hydroxybenzoic acid is predominantly measured in the surface samples suggests that exposure to light may play a role. In particular, the time of day when the sample was collected might impact whether 4-hydroxybenzoic acid was measured if it is generated through photo-oxidation. All stations where 4-hydroxybenzoic acid was measured were sampled during daylight hours with two exceptions. Stn A11 was sampled approximately 2 hours after sunset and 4-hydroxybenzoic acid was measured. Conversely, Stn K5 was sampled mid-morning but no 4-hydroxybenzoic acid was measured. While the residence time of 4-hydroxybenzoic acid in seawater is not known, to our knowledge, it could in theory persist in surface waters for several hours after dark, and take several hours of light exposure before it began to accumulate again when it was light. Alternatively, 4-hydroxybenzoic acid could be released by microbes that are dominant in the surface relative to the DCM. For example, in the

surface water cyanobacteria are particularly prevalent. In one case, *Synechococcus elongatus* CCMP 1631 released 4-hydroxybenzoic acid in culture experiments and perhaps other cyanobacteria do as well (Fiore et al. 2015). Other marine bacteria, specifically a *Microbulbifer* species isolated near Palau, also produce extracellular 4-hydroxybenzoic acid (Peng et al. 2006). These examples illustrate that metabolite concentrations are not simply a function of biomass; the abundance of some of these metabolites may be shaped by a combination of environmental factors and community composition.

4.4.2 Metabolite composition in the deep ocean

The particulate metabolites that are measured at depth (Figure 3b, group 1) indicate some novel findings as well as some trends that are supported by previous studies. They include the amino acids phenylalanine, tryptophan, (iso)leucine, proline, and arginine. Ornithine, while not one of the amino acids utilized in proteins, is an amino acid with an amine side chain but its physiological roles include being an intermediate in the urea cycle, putrescine biosynthesis, and arginine degradation as well as being a component of ornithine lipids. The two purine nucleobases guanine and adenine were also measured in a variety of deep samples. Guanine in particular was found in most deep samples (Figure 2). Interestingly, caffeine was also measured in deep samples and correlated most positively with some of the other deep metabolites (guanine, phenylalanine, tryptophan). Given our limited knowledge of caffeine in the ocean, it is not known if its presence is due to production by an organism as a defense compound or is the result of a purine degradation process. Finally, biotin and pyridoxine are both B vitamins that were measured below the euphotic zone. Biotin was only measured in a few deep samples while pyridoxine was more prevalent. It is striking that of all the B vitamins measured pyridoxine was measured at a variety of depths and with deep concentrations that were quite similar to those in

the surface suggesting that this vitamin might be of particular importance to deep sea communities. It is an essential cofactor in many enzymes associated with amino acid metabolism (Mittenhuber 2001).

The metabolites measured in the deep ocean must either be derived from its unique resident microbial community, about which relatively little is known, or be delivered from the surface ocean by rapidly sinking particles, or be products of degradation of high molecular weight organic matter associated with particles sinking from the surface ocean. Interestingly, in a study that sought to establish a degradation index for particulate organic matter based on amino acid composition, Dauwe et al. (1999) found that threonine, arginine, aspartic acid and glycine increased as a mole percent of total amino acids with increasing organic matter degradation. While we did not quantify threonine and glycine in this study, we observed arginine in many deep samples. Aspartic acid was only measured in few samples throughout the transect but some of those include deep samples. While these amino acids are used as part of a degradation index for particulate organic matter, it is unclear whether their increased proportion of deep particulate organic matter is due to *in situ* production or selective preservation on sinking particles. Tryptophan and phenylalanine are also measured throughout our deep particulate samples. However, these amino acids are not linked to organic matter degradation according to Dauwe et al. (1999), suggesting that these amino acids might be produced by the microbial community at depth.

Metabolites we might expect to measure in the deep ocean include osmolytes that are present at high concentrations in the cytosol of cells. However, betaine and glutamic acid, both common osmolytes across all the domains of life are not detected in the deep ocean. Other metabolites with similar concentrations in the euphotic zone such as guanine and phenylalanine,

respectively, are measured in the deep ocean. This suggests that betaine and glutamic acid decrease more quickly than biomass with depth. This perhaps indicates that in the deep ocean other molecules are used as osmolytes by marine microbes due to differing requirements of organisms living in high pressure environments. Deep sea invertebrates have been shown to use different osmolytes than surface organisms but, to our knowledge, this has not been examined in marine microbes (Yancey et al. 2002). The amino acid proline also functions as an osmolyte (Burg and Ferraris 2008) and was measured in many deep samples, perhaps indicating a preferential use of this molecule as an osmolyte at high hydrostatic pressure and low temperatures. It has been found to increase intracellularly with increasing osmolarity in *Bacillus subtilis* (Brill et al. 2011). Proline has not been widely characterized in studies of marine amino acid distributions, but recent work measured proline content in bacterial-sized particles and found that L-proline comprised 46% of the total (free and hydrolysable) amino acids measured in their study (Takasu and Nagata 2015). Both studies support the idea that proline may be a substantial component of cells in the deep ocean and in particular, the measurement of free proline, which functions as an osmolyte, suggests that it might be related to osmotic regulation. In the surface ocean osmolytes like DMSP are important sources of organic carbon that support heterotrophic growth, in the deep ocean the most prevalent sources of organic matter differ. For example, perhaps proline serves as an important carbon substrate for deep ocean heterotrophs.

4.4.3 Relationships between particulate and dissolved metabolites

To delve deeper into the cycling of individual metabolites, the particulate and dissolved distributions can be compared. Two B vitamins (riboflavin and pantothenic acid) are more prevalent in the dissolved fraction than in the particulate samples. These metabolites might be useful in some type of exchange between organisms as they are vital cofactors in all domains of

life. Unlike some of the other B vitamins, auxotrophy for riboflavin has not been documented to our knowledge, although it is possible that microbes would assimilate available riboflavin rather than synthesize it *de novo*. In non-marine systems riboflavin has been demonstrated to have activity beyond its role as a vitamin including acting as a quorum sensing molecule mimic (Rajamani et al. 2008) and also priming plant defense responses (Zhang et al. 2009). The versatility of this molecule suggests we may not fully understand its role in marine systems. Pantothenic acid is also synthesized by plants and microbes and is a precursor to coenzyme A and acyl carrier protein. However, to our knowledge, this vitamin has not been measured in the ocean before and its distribution and importance for marine organisms is completely unknown. It is commonly included in growth media used to cultivate marine microbes.

Of the metabolite distributions that can be compared between particulate and dissolved samples in the euphotic zone, there seem to be three types of relationships. These include 1) a linear relationship where the particulate and dissolved relative concentrations vary in tandem, 2) dissolved distributions of a metabolite that have no apparent relationship to the particulate variability, 3) a combination type distribution where in some regions of the ocean there appears to be a relationship and in other parts there is no relationship between particulate and dissolved concentrations. Tryptophan and phenylalanine are both metabolites that have a linear relationship between particulate and dissolved abundances throughout the transect in the euphotic zone (Figure 7c-d). This suggests that there is a direct relationship between production of these molecules inside the cell and their release into the water, as well as a fairly constant rate of removal.

4-Hydroxybenzoic acid, whose dissolved distribution in the euphotic zone was discussed earlier, shows no link between particulate and dissolved samples. In some cases dissolved 4-

hydroxybenzoic acid is present where particulate 4-hydroxybenzoic acid is not detected; when they are both measured, dissolved 4-hydroxybenzoic acid concentrations are similar, rather than varying with particulate concentration (Figure 7a). This lends further support to the idea, discussed earlier, that perhaps dissolved 4-hydroxybenzoic acid is not a direct product of metabolism and could be derived from photo-oxidation of organic matter in the surface water.

Finally, MTA seems to have a relationship between particulate and dissolved concentrations that varies between ocean regions. In the North Atlantic particulate and dissolved MTA is measured throughout the euphotic zone. However, in the South Atlantic, MTA is measured in most particulate samples but is not detected in most dissolved samples (Figure 7b). The latitudinal profile of particulate MTA shows a similar distribution to the other metabolites linked to biomass along the transect with the highest abundances at stations A4, A7, the DCM at K15, and the surface at K23 while also being less abundant but still elevated in most DCM samples and more so in the North Atlantic (Figure 8a). In contrast, dissolved MTA was only measured in the North Atlantic and at the surface in K23 (Amazon Plume station) where it was also high in the particulate (Figure 8b). Of particular note is that the DCM at K15 has the second highest abundance of particulate MTA, but no MTA was measured in the dissolved sample.

This difference in MTA distributions suggests that the particulate concentration of MTA does not directly determine the dissolved MTA concentration as we see for tryptophan and phenylalanine, at least not in all regions of the ocean. There are two possible causes of this mismatch: (a) that organisms are not releasing MTA from the cell in the South Atlantic samples or (b) that dissolved MTA is cycled far more rapidly in the South Atlantic through either an abiotic transformation or uptake by another organism. MTA is a byproduct of two biosynthetic reactions and it can either be recycled through the methionine salvage pathway or simply

released into the environment.

Accumulation of too much MTA inside the cell results in inhibition of the enzymes that produce it, so concentrations inside cells are maintained at relatively constant concentrations (Raina et al. 1982; Parsek et al. 1999). This is consistent with our observations in the particulate samples. Release of MTA varies by organism, however. In culture experiments the cyanobacterium, *Synechococcus elongatus*, does not release MTA (Fiore et al. 2015) while the heterotrophic bacterium, *Ruegeria pomeroyi*, does (Johnson et al. 2016). So, the contrast in dissolved MTA distribution between the North and South Atlantic could be caused by the microbial community composition associated with distinct oceanic regions. The South Atlantic oligotrophic gyre has a community dominated by cyanobacteria, particularly *Prochlorococcus* (Zubkov et al. 1998), while in the North Atlantic *Synechococcus* and picoeukaryotes are more prevalent (Li and Harrison 2001). Community differences in release of MTA could potentially affect dissolved MTA concentrations, although data on MTA production and release in more species is necessary to support this contention. Identifying this type of variation is key to understanding how different microbial communities shape dissolved organic matter composition and subsequently the community itself. MTA could be a valuable source of organic sulfur and nitrogen in the ocean and variability in its distribution could have significance for microbial community function.

4.5 Conclusions

By simultaneously measuring a suite of metabolites in a variety of oceanic regions, we can begin to assemble an understanding of metabolite variability in the ocean. We find that there is a group of metabolites that remain constant relative to biomass, while other metabolites show distinct distributions. These metabolites that deviate from biomass include nucleic acids, B

vitamins, amino acids, and a variety of metabolic intermediates. While, in many cases, these metabolites have been studied in a biomedical context or in a model organism and their function in the cell is known, surprisingly little is known about how these molecules respond to environmental cues or their specificity to certain species (particularly in the ocean). This creates challenges in interpreting these data in the environment. There are surprises, such as the finding that adenine is distributed differently than other nucleobases, that caffeine is measured in the open ocean, and that proline is a prevalent amino acid in the deep sea. This dataset provides a context for future marine metabolomics work. Linking metabolomics datasets to community composition, nutrient availability, and other omics datasets will allow further testing of hypotheses proposed here. In addition, controlled culture experiments can continue to expand our understanding of how metabolites respond to the environment and additional field studies will complete the picture of metabolite distributions in other ocean regions. As this understanding is expanded, the mechanisms that control the flux of these metabolites through the marine food web and carbon cycle may be clarified and more accurately predicted.

Acknowledgments

Funding for this work came from a National Science Foundation grant (Grant OCE-1154320 to EBK and KL). The instruments in the WHOI FT-MS Facility were purchased with support from the GBMF and NSF. Support for WMJ was provided by a National Defense Science and Engineering Fellowship. I would like to thank Melissa Kido Soule, Krista Longnecker, and the captains and crew of the R/V *Knorr* and the R/V *Atlantic Explorer*, Mike Lomas, Catherine Carmichael, Gwenn Hennon, and Erin Eggleston for helping to make this dataset possible.

References

- Aluwihare, L., D. Repeta, and R. Chen. 1997. A major biopolymeric component to dissolved organic carbon in surface sea water. *Nature* **387**: 166–169.
- Ankrah, N. Y. D., A. L. May, J. L. Middleton, and others. 2014. Phage infection of an environmentally relevant marine bacterium alters host metabolism and lysate composition. *ISME J.* **8**: 1089–1100. doi:10.1038/ismej.2013.216
- Arar, E. J., and G. B. Collins. 1997. Method 445.0: *In vitro* determination of chlorophyll *a* and pheophytin *a* in marine and freshwater algae by fluorescence.
- Bird, D. F., and J. Kalff. 1984. Empirical relationships between bacterial abundance and chlorophyll concentration in fresh and marine waters. *Can. J. Fish. Aquat. Sci.* **41**: 1015–1023. doi:10.1139/f84-118
- Brauer, M. J., J. Yuan, B. D. Bennett, W. Lu, E. Kimball, D. Botstein, and J. D. Rabinowitz. 2006. Conservation of the metabolomic response to starvation across two divergent microbes. *Proc. Natl. Acad. Sci. U. S. A.* **103**: 19302–19307. doi:10.1073/pnas.0609508103
- Brill, J., T. Hoffmann, M. Bleisteiner, and E. Bremer. 2011. Osmotically controlled synthesis of the compatible solute proline is critical for cellular defense of *Bacillus subtilis* against high osmolarity. *J. Bacteriol.* **193**: 5335–5346. doi:10.1128/JB.05490-11
- Burg, M. B., and J. D. Ferraris. 2008. Intracellular organic osmolytes: Function and regulation. *J. Biol. Chem.* **283**: 7309–7313. doi:10.1074/jbc.R700042200
- Carini, P., E. O. Campbell, J. Morré, and others. 2014. Discovery of a SAR11 growth requirement for thiamin's pyrimidine precursor and its distribution in the Sargasso Sea. *ISME J.* **8**: 1727–38. doi:10.1038/ismej.2014.61
- Carlson, C. A., and D. A. Hansell. 2015. DOM sources, sinks, reactivity, and budgets, p. 65–102. *In* D.A. Hansell and C.A. Carlson [eds.], *Biogeochemistry of marine dissolved organic matter*. Elsevier.
- Chambers, M. C., B. Maclean, R. Burke, and others. 2012. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**: 918–920. doi:10.1038/nbt.2377
- Clasquin, M. F., E. Melamud, and J. D. Rabinowitz. 2012. LC-MS data processing with MAVEN: A metabolomic analysis and visualization engine. *Curr. Protoc. Bioinforma.* **37**: 1–23. doi:10.1002/0471250953.bi1411s37
- Croft, M. T., M. J. Warren, and A. G. Smith. 2006. Algae need their vitamins. *Eukaryot. Cell* **5**: 1175–1183. doi:10.1128/EC.00097-06
- Dauwe, B., J. J. Middelburg, P. M. J. Herman, and C. H. R. Heip. 1999. Linking diagenetic alteration of amino acids and bulk organic matter reactivity. *Limnol. Oceanogr.* **44**: 1809–1814. doi:10.4319/lo.1999.44.7.1809
- Dittmar, T., H. Fitznar, and G. Kattner. 2001. Origin and biogeochemical cycling of organic nitrogen in the eastern Arctic Ocean as evident from D- and L-amino acids. *Geochim. Cosmochim. Acta* **65**: 4103–4114.
- Dittmar, T., B. Koch, N. Hertkorn, and G. Kattner. 2008. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol. Oceanogr. Methods* **6**: 230–235. doi:10.4319/lom.2008.6.230
- Dittmar, T., and B. P. Koch. 2006. Thermogenic organic matter dissolved in the abyssal ocean. *Mar. Chem.* **102**: 208–217. doi:10.1016/j.marchem.2006.04.003
- Ducklow, H. W. 1999. Minireview: The bacterial content of the oceanic euphotic zone. *FEMS Microbiol.* **30**: 1–10. doi:10.1016/S0168-6496(99)00031-8

- Fiehn, O. 2002. Metabolomics—the link between genotypes and phenotypes. *Plant Mol. Biol.* **48**: 155–71.
- Fiore, C. L., K. Longnecker, M. C. Kido Soule, and E. B. Kujawinski. 2015. Release of ecologically relevant metabolites by the cyanobacterium, *Synechococcus elongatus* CCMP 1631. *Environ. Microbiol.* **17**: 3949–3963. doi:10.1111/1462-2920.12899
- Gebser, B., and G. Pohnert. 2013. Synchronized regulation of different zwitterionic metabolites in the osmoadaptation of phytoplankton. *Mar. Drugs* **11**: 2168–82. doi:10.3390/md11062168
- Hedges, J. I., and P. L. Parker. 1976. Land-derived organic matter in surface sediments from the Gulf of Mexico. *Geochim. Cosmochim. Acta* **40**: 1019–1029. doi:10.1016/0016-7037(76)90044-2
- Hertkorn, N., R. Benner, M. Frommberger, P. Schmitt-Kopplin, M. Witt, K. Kaiser, A. Kettrup, and J. I. Hedges. 2006. Characterization of a major refractory component of marine dissolved organic matter. *Geochim. Cosmochim. Acta* **70**: 2990–3010. doi:10.1016/j.gca.2006.03.021
- Johnson, W. M., M. C. Kido Soule, and E. B. Kujawinski. Extraction efficiency and quantification of dissolved metabolites in targeted marine metabolomics. submitted.
- Johnson, W. M., M. C. Kido Soule, and E. B. Kujawinski. 2016. Evidence for quorum sensing and differential metabolite production by the marine heterotroph, *Ruegeria pomeroyi*, in response to DMSP. *ISME J.* **10**: 2304–2316. doi:10.1038/ismej.2016.6
- Kaiser, K., and R. Benner. 2009. Biochemical composition and size distribution of organic matter at the Pacific and Atlantic time-series stations. *Mar. Chem.* **113**: 63–77. doi:10.1016/j.marchem.2008.12.004
- Kido Soule, M. C., K. Longnecker, W. M. Johnson, and E. B. Kujawinski. 2015. Environmental metabolomics: Analytical strategies. *Mar. Chem.* **177**: 374–387. doi:10.1016/j.marchem.2015.06.029
- Kiene, R. P., L. J. Linn, and J. a. Bruton. 2000. New and important roles for DMSP in marine microbial communities. *J. Sea Res.* **43**: 209–224. doi:10.1016/S1385-1101(00)00023-X
- Kruskopf, M., and K. J. Flynn. 2006. Chlorophyll content and fluorescence responses cannot be used to gauge reliably phytoplankton biomass, nutrient status or growth rate. *New Phytol.* **169**: 525–536. doi:10.1111/j.1469-8137.2005.01601.x
- Lee, C., and C. Cronin. 1982. The vertical flux of particulate organic nitrogen in the sea: decomposition of amino acids in the Peru upwelling area and the equatorial Atlantic. *J. Mar. Res.* **40**: 227–251.
- Lee, C., and C. Cronin. 1984. Particulate amino acids in the sea: Effects of primary productivity and biological decomposition. *J. Mar. Res.* **42**: 1075–1097. doi:10.1357/002224084788520710
- Lesser, M. P. 2006. Oxidative stress in marine environments: Biochemistry and physiological ecology. *Annu. Rev. Physiol.* **68**: 253–278. doi:10.1146/annurev.physiol.68.040104.110001
- Li, W. K. W., and W. G. Harrison. 2001. Chlorophyll, bacteria and picophytoplankton in ecological provinces of the North Atlantic. *Deep. Res. Part II Top. Stud. Oceanogr.* **48**: 2271–2293. doi:10.1016/S0967-0645(00)00180-6
- Longnecker, K., M. C. Kido Soule, and E. B. Kujawinski. 2015. Dissolved organic matter produced by *Thalassiosira pseudonana*. *Mar. Chem.* **168**: 114–123. doi:10.1016/j.marchem.2014.11.003
- McCarthy, M., J. Hedges, and R. Benner. 1996. Major biochemical composition of dissolved high molecular weight organic matter in seawater. *Mar. Chem.* **55**: 281–297.

doi:10.1016/S0304-4203(96)00041-2

- Melamud, E., L. Vastag, and J. D. Rabinowitz. 2010. Metabolomic analysis and visualization engine for LC-MS data. *Anal. Chem.* **82**: 9818–9826.
- Mithöfer, A., and W. Boland. 2012. Plant defense against herbivores: Chemical aspects. *Annu. Rev. Plant Biol.* **63**: 431–450. doi:10.1146/annurev-arplant-042110-103854
- Mittenhuber, G. 2001. Phylogenetic analyses and comparative genomics of vitamin B6 (pyridoxine) and pyridoxal phosphate biosynthesis pathways. *J. Mol. Microbiol. Biotechnol.* **3**: 1–20.
- Mopper, K., and P. Lindroth. 1982. Diel and depth variations in dissolved free amino acids and ammonium in the Baltic Sea determined by shipboard HPLC analysis. *Limnol. Oceanogr.* **27**: 336–347. doi:10.4319/lo.1982.27.2.0336
- Nagata, T. 2008. Organic matter–bacteria interactions in seawater, p. 204–241. *In* D. Kirchman [ed.], *Microbial Ecology of the Oceans*. Wiley-Blackwell.
- Noble, R. T., and J. A. Fuhrman. 1998. Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. *Aquat. Microb. Ecol.* **14**: 113–118. doi:10.3354/ame014113
- Page, S. E., W. A. Arnold, and K. McNeill. 2011. Assessing the contribution of free hydroxyl radical in organic matter- sensitized photo-hydroxylation reactions - Supporting Information. *Environ. Sci. Technol.* **45**: 1–11.
- Parsek, M. R., D. L. Val, B. L. Hanzelka, J. E. Cronan, and E. P. Greenberg. 1999. Acyl homoserine-lactone quorum-sensing signal generation. *Proc. Natl. Acad. Sci. U. S. A.* **96**: 4360–5. doi:10.1073/pnas.96.8.4360
- Peng, X., K. Adachi, C. Chen, H. Kasai, K. Kanoh, Y. Shizuri, and N. Misawa. 2006. Discovery of a marine bacterium producing 4-hydroxybenzoate and its alkyl esters, parabens. *Appl. Environ. Microbiol.* **72**: 5556–5561. doi:10.1128/AEM.00494-06
- Pohnert, G. 2000. Wound-activated chemical defense in a unicellular planktonic algae. *Angew. Chemie Int. Ed.* **30**: 4352–4354.
- Rabinowitz, J. D., and E. Kimball. 2007. Acidic acetonitrile for cellular metabolome extraction from *Escherichia coli*. *Anal. Chem.* **79**: 6167–6173.
- R Core Team. 2015. R: A language and environment for statistical computing.
- Raina, A., K. Tuomi, and R. L. Pajula. 1982. Inhibition of the synthesis of polyamines and macromolecules by 5'-methylthioadenosine and 5'-alkylthiotubercidins in BHK21 cells. *Biochem. J.* **204**: 697–703.
- Rajamani, S., W. D. Bauer, J. B. Robinson, and others. 2008. The vitamin riboflavin and its derivative lumichrome activate the LasR bacterial quorum-sensing receptor. *Mol. plant-microbe Interact.* **21**: 1184–92. doi:10.1094/MPMI-21-9-1184
- Rich, J., M. Gosselin, E. Sherr, B. Sherr, and D. L. Kirchman. 1997. High bacterial production, uptake and concentrations of dissolved organic matter in the Central Arctic Ocean. *Deep. Res. Part II Top. Stud. Oceanogr.* **44**: 1645–1663. doi:10.1016/S0967-0645(97)00058-1
- Rinas, U., K. Hellmuth, R. Kang, A. Seeger, and H. Schlieker. 1995. Entry of *Escherichia coli* into stationary phase is indicated by endogenous and exogenous accumulation of nucleobases. *Appl. Environ. Microbiol.* **61**: 4147–4151.
- Rodriguez del Rey, Z., E. F. Granek, and S. Sylvester. 2012. Occurrence and concentration of caffeine in Oregon coastal waters. *Mar. Pollut. Bull.* **64**: 1417–1424. doi:10.1016/j.marpolbul.2012.04.015
- Sañudo-Wilhelmy, S. A., L. S. Cutter, R. Durazo, and others. 2012. Multiple B-vitamin depletion in large areas of the coastal ocean. *Proc. Natl. Acad. Sci. U. S. A.* **109**: 14041–5.

- doi:10.1073/pnas.1208755109
- Schlitzer, R. 2016. Ocean Data View.
- Seyedsayamdost, M. R., R. J. Case, R. Kolter, and J. Clardy. 2011. The Jekyll-and-Hyde chemistry of *Phaeobacter gallaeciensis*. *Nat. Chem.* **3**: 331–335. doi:10.1038/nchem.1002
- Søndergaard, M., P. J. le B. Williams, G. Cauvet, B. Riemann, C. Robinson, S. Terzic, E. M. S. Woodward, and J. Worm. 2000. Net accumulation and flux of dissolved organic carbon and dissolved organic nitrogen in marine plankton communities. *Limnol. Oceanogr.* **45**: 1097–1111.
- Takasu, H., and T. Nagata. 2015. High proline content of bacteria-sized particles in the Western North Pacific and its potential as a new biogeochemical indicator of organic matter diagenesis. *Front. Mar. Sci.* **2**: 110. doi:10.3389/fmars.2015.00110
- Thompson, P. A., M. X. Guo, P. J. Harrison, and J. N. C. Whyte. 1992. Effects of variation in temperature. II. On the fatty-acid composition of eight species of marine phytoplankton. *J. Phycol.* **28**: 488–497. doi:10.1111/j.0022-3646.1992.00488.x
- Vione, D., M. Ponzo, D. Bagnus, V. Maurino, C. Minero, and M. E. Carlotti. 2010. Comparison of different probe molecules for the quantification of hydroxyl radicals in aqueous solution. *Environ. Chem. Lett.* **8**: 95–100. doi:10.1007/s10311-008-0197-3
- Weigel, S., J. Kuhlmann, and H. Hühnerfuss. 2002. Drugs and personal care products as ubiquitous pollutants: Occurrence and distribution of clofibric acid, caffeine and DEET in the North Sea. *Sci. Total Environ.* **295**: 131–141. doi:10.1016/S0048-9697(02)00064-5
- Winkler, W. C., and R. R. Breaker. 2005. Regulation of Bacterial Gene Expression By Riboswitches. *Annu. Rev. Microbiol.* **59**: 487–517. doi:doi:10.1146/annurev.micro.59.030804.121336
- Yancey, P. H., W. R. Blake, and J. Conley. 2002. Unusual organic osmolytes in deep-sea animals: Adaptations to hydrostatic pressure and other perturbants. *Comp. Biochem. Physiol. - A Mol. Integr. Physiol.* **133**: 667–676. doi:10.1016/S1095-6433(02)00182-4
- Yentsch, C. S., and D. W. Menzel. 1963. A method for the determination of phytoplankton chlorophyll and phaeophytin by fluorescence. *Deep Sea Res. Oceanogr. Abstr.* **10**: 221–231. doi:10.1016/0011-7471(63)90358-9
- Zhang, S., X. Yang, M. Sun, F. Sun, S. Deng, and H. Dong. 2009. Riboflavin-induced priming for pathogen defense in *Arabidopsis thaliana*. *J. Integr. Plant Biol.* **51**: 167–174. doi:10.1111/j.1744-7909.2008.00763.x
- Zubkov, M. V., M. A. Sleigh, G. A. Tarran, P. H. Burkill, and R. J. G. Leakey. 1998. Picoplanktonic community structure on an Atlantic transect from 50°N to 50°S. *Deep. Res. Part I Oceanogr. Res. Pap.* **45**: 1339–1355. doi:10.1016/S0967-0637(98)00015-6

Table 1. Range of particulate concentrations of each metabolite and prevalence in samples. Minimum concentration is the lowest value in a sample where the metabolite was detected. ^aPicomoles of particulate metabolite per liter filtered. Values with a “<” symbol indicate the detection limit of the metabolite as the metabolite was not detected in any sample. n.d. indicates that the LOD is not determined.

| Compound | Min ^a pM (>0) | Max ^a pM | # of samples <200m (n=28) | # of samples >200m (n=46) |
|----------------------------------|--------------------------------|------------------------|---------------------------------|---------------------------------|
| 1-deoxy-D-xylulose-5-phosphate | 0.1 | 0.8 | 4 | 3 |
| 2,3-dihydroxybenzoic acid | 0.9 | 0.9 | 1 | 0 |
| 2,3-dihydroxypropane-1-sulfonate | 8.7 | 72 | 4 | 3 |
| 4-hydroxybenzoic acid | 0.6 | 2.9 | 2 | 2 |
| 5-methylthioadenosine | 0.1 | 1.1 | 23 | 2 |
| adenine | 0.3 | 10 | 11 | 21 |
| adenosine | 0.08 | 77 | 23 | 17 |
| adenosine 5'-monophosphate | 0.02 | 27 | 26 | 25 |
| arginine | 0.4 | 7.8 | 9 | 22 |
| aspartic acid | 55 | 320 | 8 | 5 |
| betaine | 3.8 | 580 | 27 | 7 |
| biotin | 0.03 | 1.0 | 4 | 5 |
| caffeine | 0.02 | 4.2 | 9 | 34 |
| chitobiose | 1.3 | 5.3 | 5 | 0 |
| ciliatine | 0.1 | 0.3 | 1 | 3 |
| citrulline | 0.3 | 9.1 | 8 | 0 |
| desthiobiotin | 0.3 | 0.3 | 1 | 0 |
| D-glucosamine 6-phosphate | 0.06 | 0.1 | 0 | 2 |
| dihydroxyacetone phosphate | 0.2 | 1.8 | 1 | 2 |
| DMSP | 7.6 | 3700 | 28 | 8 |
| D-ribose 5-phosphate | 0.1 | 1.7 | 3 | 2 |
| folic acid | 0.1 | 0.7 | 2 | 2 |
| glucose 6-phosphate | 0.08 | 0.7 | 2 | 4 |
| glutamic acid | 0.3 | 46 | 25 | 7 |
| glutamine | 0.7 | 3.7 | 2 | 0 |
| guanine | 0.6 | 490 | 28 | 43 |
| indole 3-acetic acid | 0.2 | 0.2 | 1 | 0 |
| inosine 5'-monophosphate | 0.2 | 0.2 | 1 | 0 |
| inosine | 0.03 | 11 | 23 | 14 |
| (iso)leucine | 0.1 | 20 | 20 | 20 |
| cysteine | 13 | 13 | 1 | 0 |
| glutathione | 7.1 | 7.1 | 1 | 0 |
| methionine | 0.3 | 8.2 | 4 | 2 |
| N-acetylglutamic acid | 0.7 | 2.0 | 3 | 0 |
| ornithine | 0.1 | 8.4 | 10 | 13 |
| orotic acid | 2.04 | 29 | 2 | 2 |

| | | | | |
|----------------------------------------|--------|--------|----|----|
| phenylalanine | 0.04 | 44 | 28 | 43 |
| proline | 0.08 | 37 | 19 | 25 |
| pyridoxine | 0.05 | 0.8 | 8 | 20 |
| riboflavin | 0.003 | 0.8 | 12 | 1 |
| <i>S</i> -(5'-adenosyl)-L-homocysteine | 1.0 | 1.0 | 1 | 0 |
| <i>sn</i> -glycerol 3-phosphate | 3.6 | 3.6 | 1 | 0 |
| threonine | 0.114 | 0.796 | 1 | 1 |
| tryptophan | 0.02 | 17 | 27 | 32 |
| uracil | 0.6 | 34 | 8 | 5 |
| uridine 5'-monophosphate | 0.07 | 0.5 | 4 | 3 |
| xanthine | 0.04 | 4.8 | 19 | 5 |
| 3-mercaptopropionic acid | < 12 | < 12 | 0 | 0 |
| 4-aminobenzoic acid | < 0.59 | < 0.59 | 0 | 0 |
| cyanocobalamin | < 0.11 | < 0.11 | 0 | 0 |
| cytosine | < 2.2 | < 2.2 | 0 | 0 |
| fosfomycin | < 53 | < 53 | 0 | 0 |
| fumaric acid | n.d. | n.d. | 0 | 0 |
| glyphosate | < 2 | < 2 | 0 | 0 |
| <i>N</i> -acetylglucosamine | < 9 | < 9 | 0 | 0 |
| pantothenic acid | < 6.7 | < 6.7 | 0 | 0 |
| sarcosine | < 48 | < 48 | 0 | 0 |
| taurine | < 0.4 | < 0.4 | 0 | 0 |
| thymidine | < 10 | < 10 | 0 | 0 |

Table 2. Range of dissolved concentrations of each metabolite and prevalence in samples. Minimum concentration is the lowest value in a sample where the metabolite was detected. Values with a “<” symbol indicate the detection limit of the metabolite as the metabolite was not detected in any sample.

| Compound | Min (pM) (>0) | Max (pM) | # of samples <200m (n=26) | # of samples >200m (n=28) |
|----------------------------------------|---------------------|-------------|---------------------------------|---------------------------------|
| 2,3-dihydroxybenzoic acid | 4.2 | 9.9 | 3 | 0 |
| 4-aminobenzoic acid | 16 | 51 | 25 | 0 |
| 4-hydroxybenzoic acid | 15 | 61 | 7 | 0 |
| 5-methylthioadenosine | 1.1 | 2.6 | 9 | 0 |
| biotin | 3.5 | 26 | 6 | 2 |
| caffeine | 2.6 | 8.7 | 12 | 4 |
| chitotriose | 12 | 130 | 7 | 1 |
| desthiobiotin | 27 | 27 | 1 | 0 |
| indole 3-acetic acid | 5.7 | 5.7 | 1 | 0 |
| inosine | 16 | 29 | 2 | 0 |
| (iso)leucine | 90 | 230 | 4 | 0 |
| <i>N</i> -acetylglutamic acid | 220 | 600 | 1 | 1 |
| NAD | 56 | 56 | 1 | 0 |
| pantothenic acid | 1.0 | 19 | 22 | 0 |
| phenylalanine | 2.2 | 290 | 26 | 12 |
| riboflavin | 0.7 | 12 | 10 | 7 |
| thiamin | 4000 | 4000 | 1 | 0 |
| tryptophan | 20 | 190 | 18 | 1 |
| 3-mercaptopropionic acid | < 78 | < 78 | 0 | 0 |
| adenosine | < 123 | < 123 | 0 | 0 |
| cyanocobalamin | < 0.8 | < 0.8 | 0 | 0 |
| folic acid | < 0.7 | < 0.7 | 0 | 0 |
| pyridoxine | < 236 | < 236 | 0 | 0 |
| <i>S</i> -(5'-adenosyl)-L-homocysteine | < 72 | < 72 | 0 | 0 |
| taurocholic acid | < 5 | < 5 | 0 | 0 |
| thymidine | < 123 | < 123 | 0 | 0 |

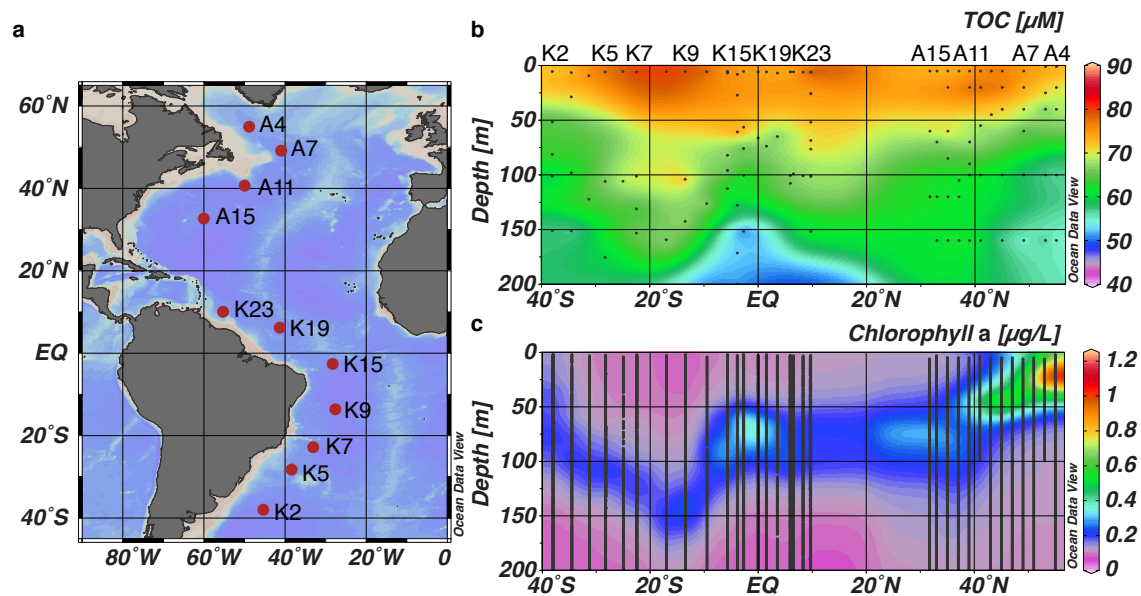


Figure 1. Transect stations and chemical parameters. (a) Stations (red dots) sampled along a latitudinal transect in the western Atlantic Ocean. (b) The total organic carbon (TOC) concentration in the top 200 m along the transect. Black dots indicate sample locations. (c) Chlorophyll *a* concentrations in the top 200 m along the transect. Lines indicate where fluorescence was measured.

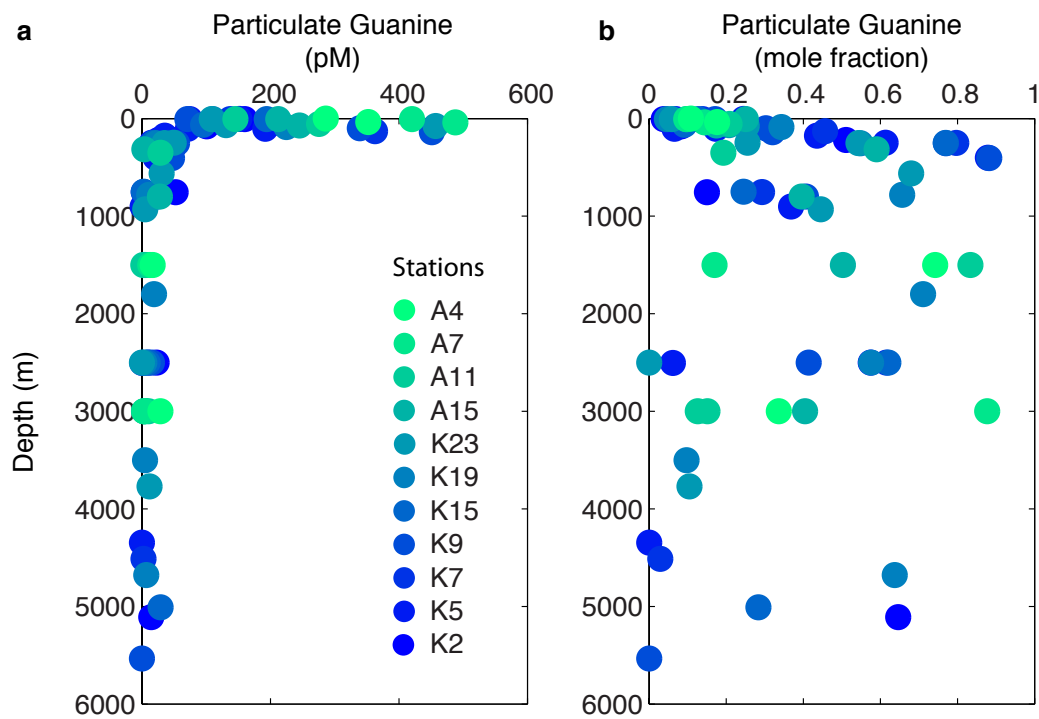


Figure 2. Particulate guanine abundance with depth. (a) Picomoles particulate guanine / liter of seawater sampled, profiled by depth for each station. (b) Moles of guanine / total moles targeted metabolites, profiled by depth for each station. Stations are indicated by different symbol colors, as shown in the legend.

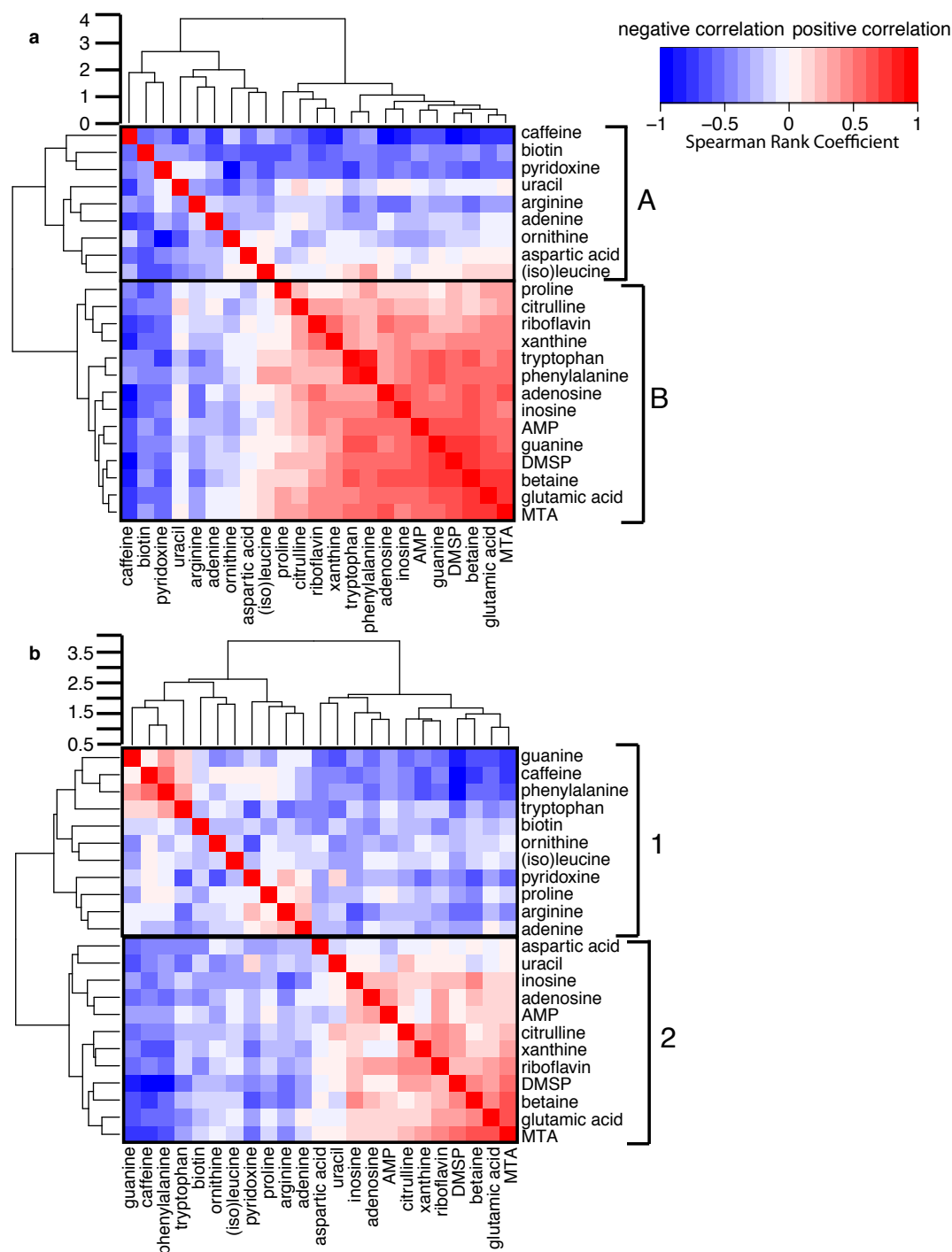


Figure 3. Spearman rank correlation coefficients plotted as a heatmap with hierarchical clustering. Only metabolites detected in at least 10% of samples are shown. Red indicates a positive correlation and blue indicates a negative correlation. (a) Comparison of particulate metabolite (pM) distributions. (b) Comparison of particulate metabolite (mole fraction) distributions. Boxes indicate metabolites within the two clusters separated by the greatest distance (see height indicated on the top dendrogram) and are referred to in the text.

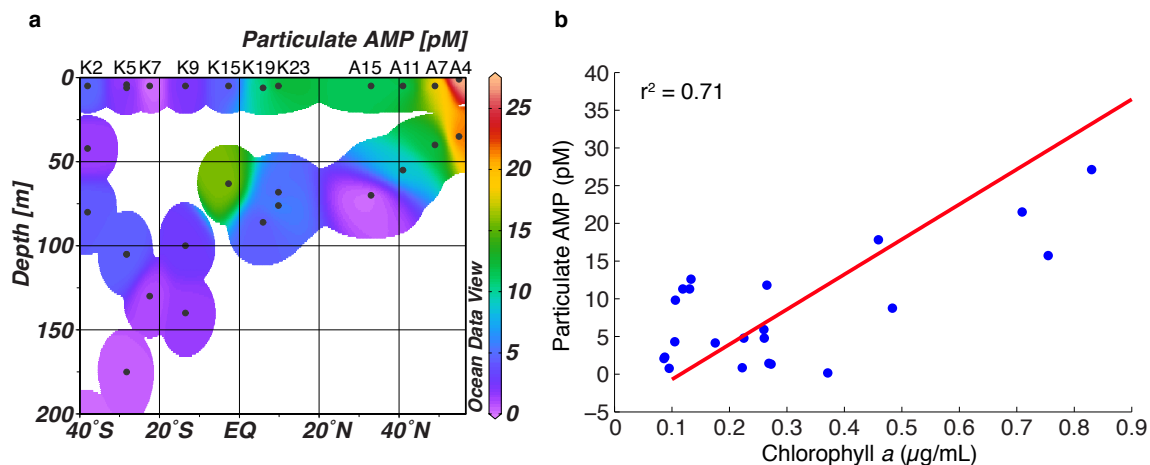


Figure 4. Adenosine 5'-monophosphate (AMP) abundance with depth (a) and relative to chlorophyll *a* concentrations (b). (a) AMP abundance (pM) above 200 m. (b) Linear regression of chlorophyll *a* versus AMP, with r^2 value (slope = 46.4, y-intercept = -5.3, standard deviation of slope = 9.8, standard deviation of intercept = 5.5).

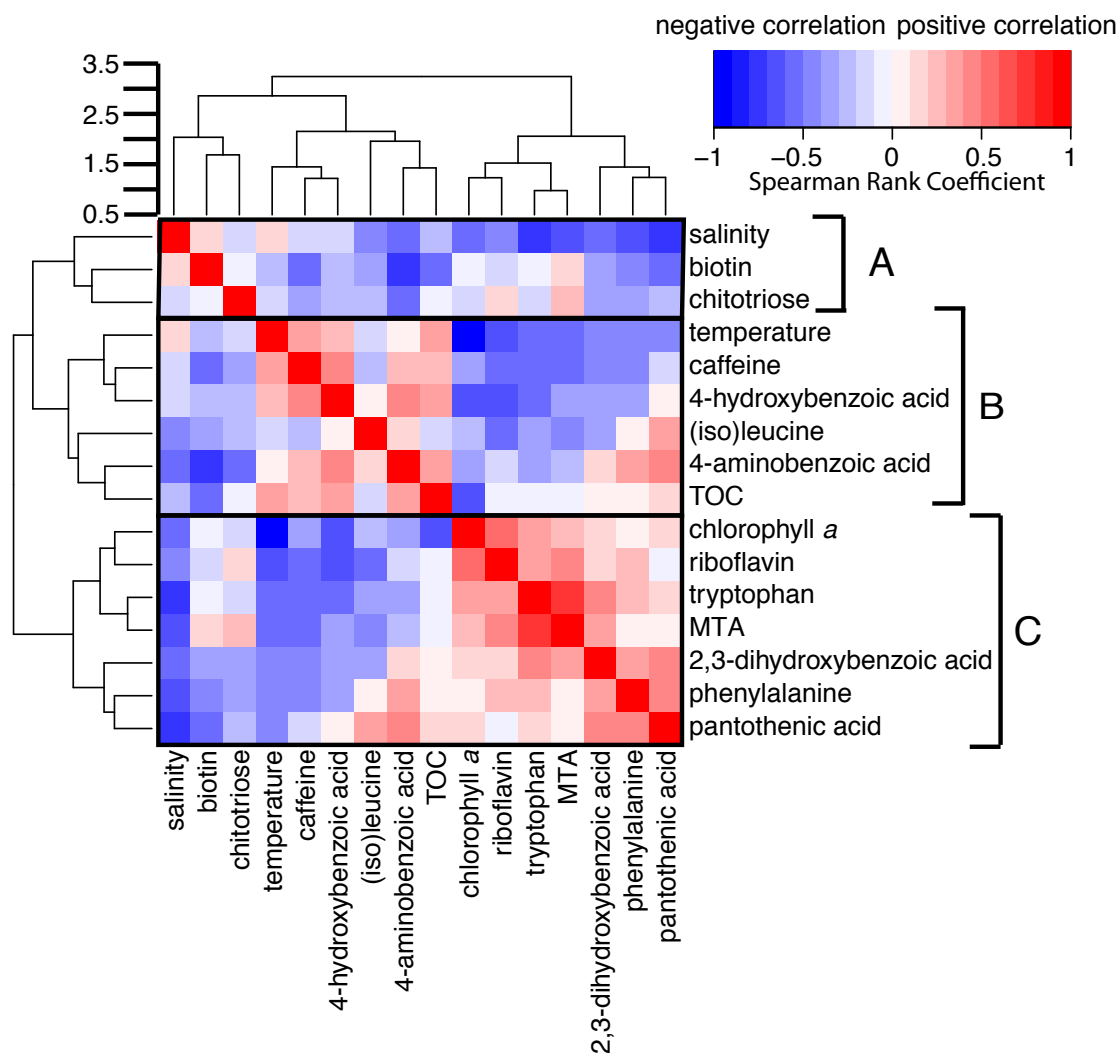


Figure 5. Spearman rank correlation of the distribution of dissolved metabolites (pM) and metadata (temperature, salinity, TOC, chlorophyll *a*) above 200 m. Only metabolites detected in at least 10% of samples are shown. Red indicates a positive correlation and blue indicates a negative correlation. Boxes indicate metabolites within the three clusters separated by the greatest distance (see height indicated on the top dendrogram) and are referred to in the text.

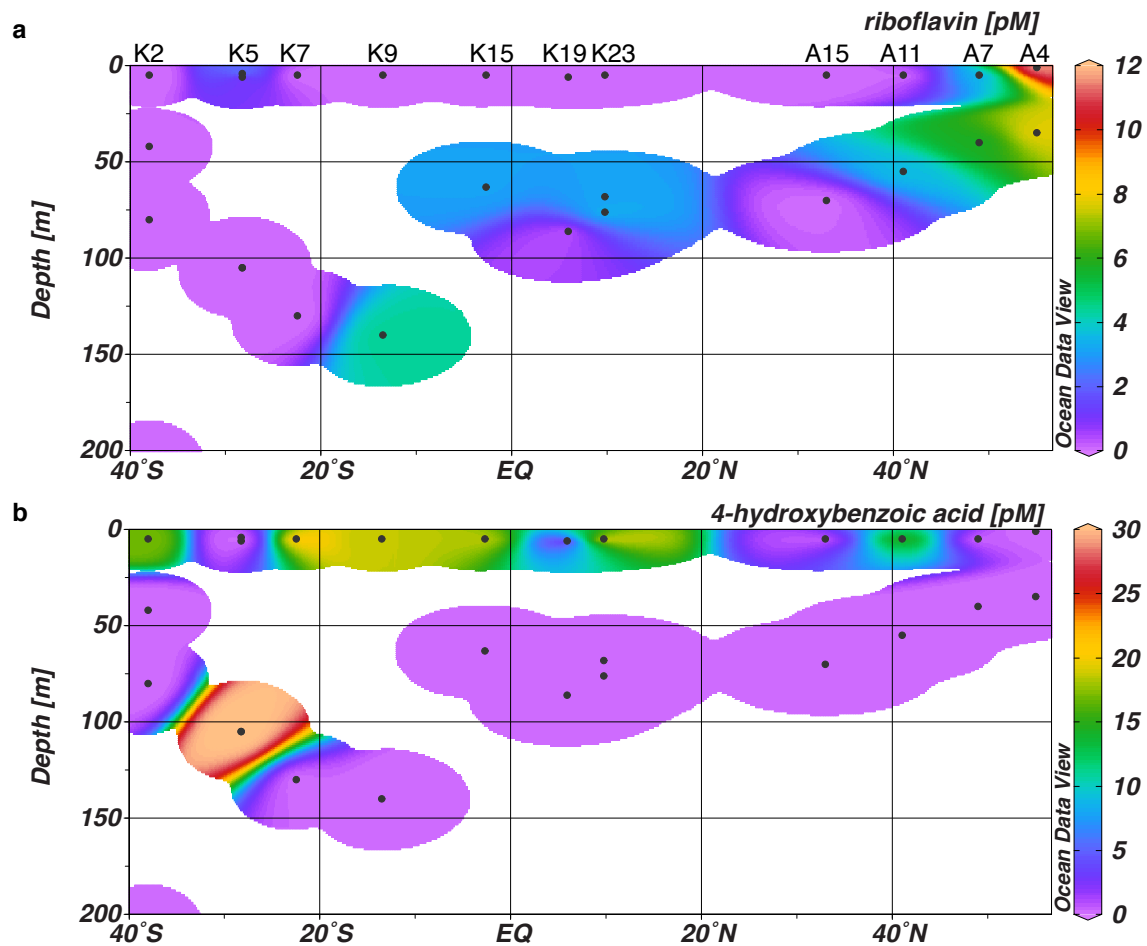


Figure 6. Concentrations of dissolved riboflavin and 4-hydroxybenzoic acid (pM). (a) Dissolved riboflavin (pM) in the top 200 m of the water column. (b) Dissolved 4-hydroxybenzoic acid (pM) in the top 200 m.

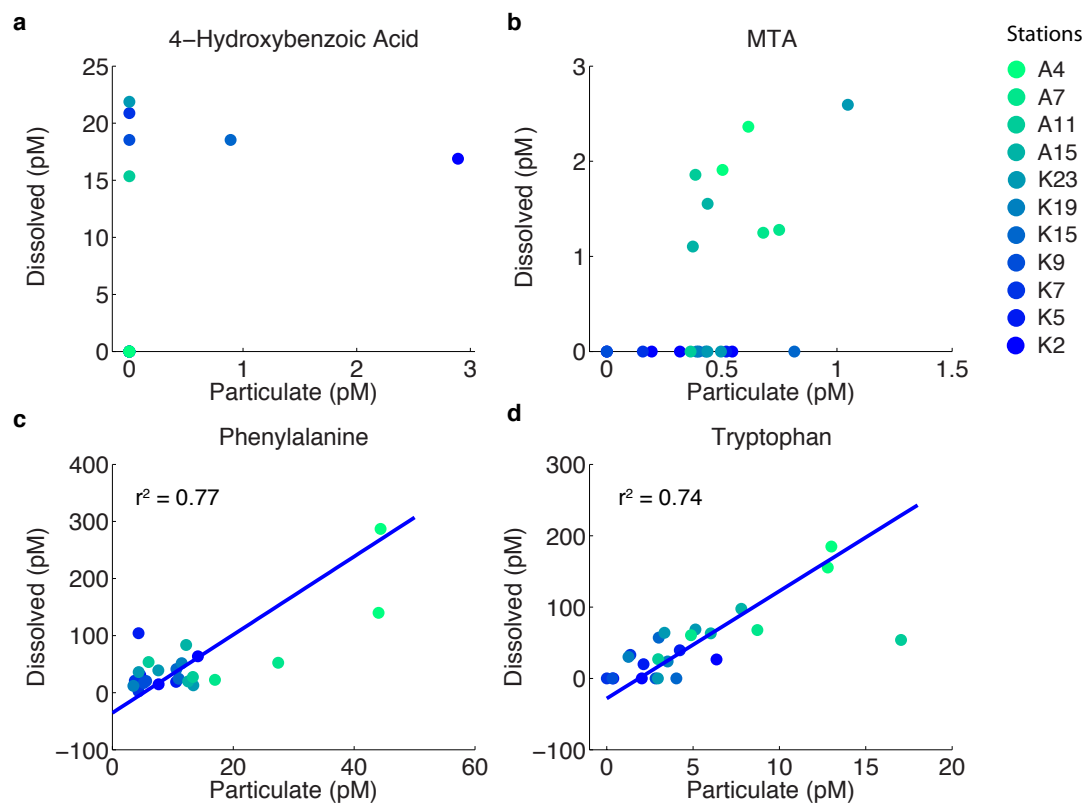


Figure 7. Comparison of particulate and dissolved metabolite distributions above 200 m. Colors indicate latitude of each sample and all samples are in moles normalized to the volume filtered. (a) 4-Hydroxybenzoic acid, (b) MTA, (c) Phenylalanine, (d) Tryptophan.

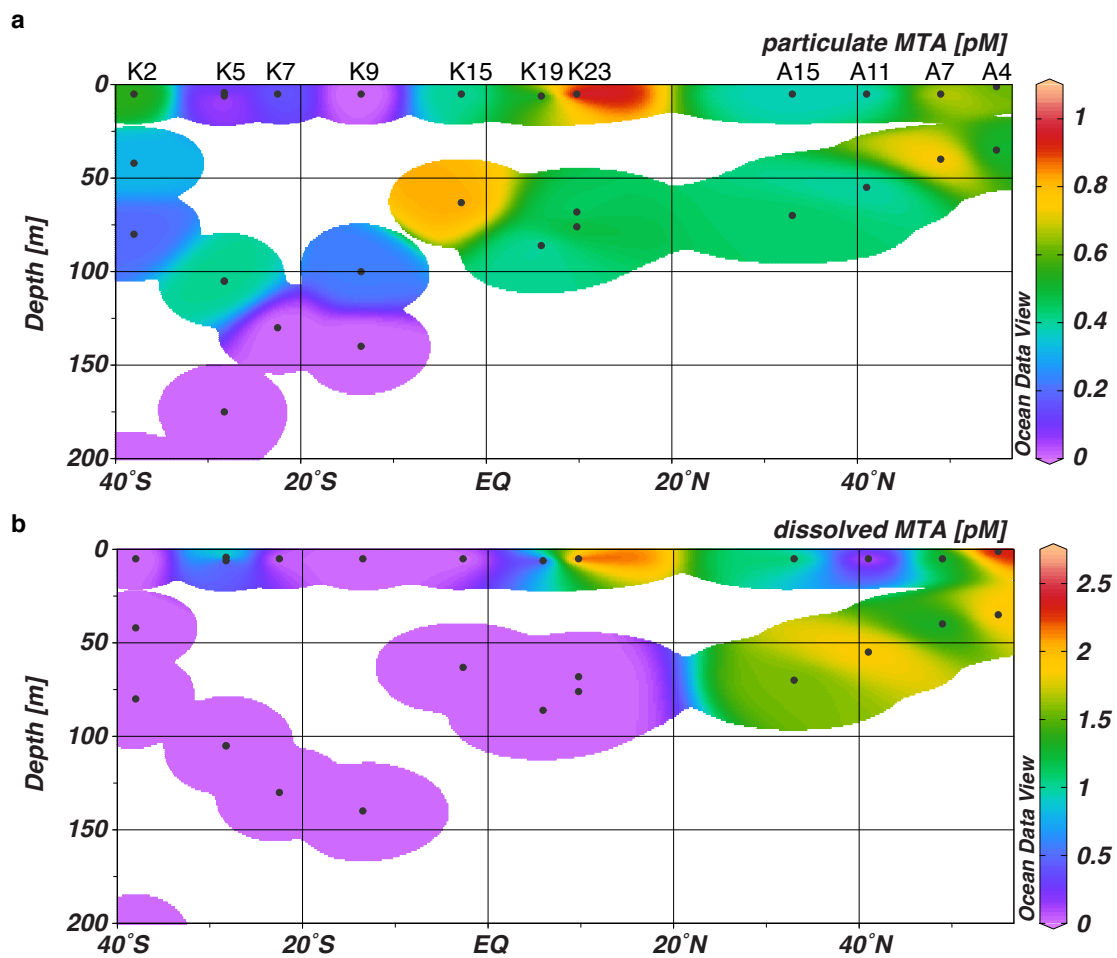


Figure 8. Dissolved and particulate 5-methylthioadenosine (MTA) abundance (pM) across transect. (a) Particulate MTA normalized to the total volume filtered. (b) pM dissolved MTA.

Supplemental Information

Table S1. Dissolved sample dilutions. Due to significant ion suppression in the middle of the chromatogram certain samples were diluted, mostly samples below the euphotic zone.

| Station | Depth | Dilution |
|---------|-------|-------------|
| K2 | 5 | no dilution |
| K2 | 42 | no dilution |
| K2 | 80 | no dilution |
| K2 | 220 | no dilution |
| K2 | 754 | 2-fold |
| K2 | 2502 | 5-fold |
| K2 | 5110 | 2-fold |
| K7 | 5 | no dilution |
| K7 | 130 | no dilution |
| K7 | 250 | no dilution |
| K7 | 753 | 5-fold |
| K7 | 2500 | 2-fold |
| K7 | 4512 | 5-fold |
| K15 | 5 | no dilution |
| K15 | 63 | no dilution |
| K15 | 250 | no dilution |
| K15 | 750 | 2-fold |
| K15 | 2500 | no dilution |
| K15 | 5009 | no dilution |
| K23 | 5 | no dilution |
| K23 | 76 | no dilution |
| K23 | 68 | no dilution |
| K23 | 251 | no dilution |
| K23 | 248 | no dilution |
| K23 | 562 | 5-fold |
| K23 | 930 | 2-fold |
| K23 | 2500 | 2-fold |
| K23 | 3770 | 2-fold |
| K5 | 4 | no dilution |
| K5 | 6 | no dilution |
| K5 | 105 | no dilution |
| K5 | 105 | no dilution |
| K5 | 248 | no dilution |
| K9 | 5 | no dilution |
| K9 | 140 | no dilution |
| K9 | 253 | no dilution |
| K19 | 6 | no dilution |
| K19 | 86 | no dilution |
| K19 | 252 | no dilution |

| | | |
|-----|------|-------------|
| A4 | 1 | no dilution |
| A4 | 35 | no dilution |
| A4 | 1500 | 2-fold |
| A4 | 3000 | 2-fold |
| A7 | 5 | no dilution |
| A7 | 40 | no dilution |
| A11 | 5 | no dilution |
| A11 | 55 | no dilution |
| A11 | 350 | 2-fold |
| A15 | 5 | no dilution |
| A15 | 70 | no dilution |
| A15 | 315 | 2-fold |
| A15 | 800 | 2-fold |
| A15 | 1500 | 2-fold |
| A15 | 3000 | 2-fold |

Table S2. Estimates of residence time, in days, assuming cell turnover rates of approximately 2 days (Ducklow 2000). The minimum, maximum, and average residence times of the samples where both the particulate and dissolved metabolite had been detected are reported. The final column indicates the number of samples for which the estimated residence time could be calculated according to the equation:

$$r = \frac{[C]_d}{[C]_p/t}$$

where r is the residence time, $[C]_d$ is the dissolved concentration of the metabolite, $[C]_p$ is the particulate concentration of the metabolite, and t is the turnover rate of cells in days, in this case estimated as 2.

| Metabolite | Min (d) | Max (d) | Mean (d) | # of samples |
|---------------------------|---------|---------|----------|--------------|
| 2,3-dihydroxybenzoic acid | 21 | 21 | 21 | 1 |
| 4-hydroxybenzoic acid | 12 | 42 | 27 | 2 |
| 5-methylthioadenosine | 3 | 18 | 8 | 9 |
| biotin | 85 | 189 | 136 | 3 |
| caffeine | 5 | 194 | 47 | 8 |
| inosine | 3 | 6 | 5 | 2 |
| (iso)leucine | 50 | 50 | 50 | 2 |
| phenylalanine | 1 | 119 | 16 | 38 |
| riboflavin | 19 | 195 | 58 | 8 |
| tryptophan | 6 | 56 | 28 | 19 |

Table S3. Previous measurements of dissolved metabolites in the ocean (n.d. : not detected).

| Compound | Dissolved (pM) | Reference |
|----------------------------------------|--------------------|---------------------------------|
| 2,3-dihydroxybenzoic acid | Not found | |
| 3-mercaptopropionic acid | 2000-3000 ∞ | (Hu et al. 2006) |
| 4-aminobenzoic acid | Not found | |
| 4-hydroxybenzoic acid | Not found | |
| 5-methylthioadenosine | Not found | |
| adenosine | Not found | |
| biotin | 10-200* | (Sañudo-Wilhelmy et al. 2012) |
| caffeine | No non-anthro | |
| chitotriose | Not found | |
| cyanocobalamin | 2-40* | (Sañudo-Wilhelmy et al. 2012) |
| desthiobiotin | Not found | |
| folic acid | 100-4000¶ | (Le Gall and Van Den Berg 1998) |
| indole 3-acetic acid | 5-400• | (Amin et al. 2015) |
| inosine | Not found | |
| (iso)leucine | 790-8000§ | (Mopper and Lindroth 1982) |
| <i>N</i> -acetylglutamic acid | Not found | |
| NAD | Not found | |
| pantothenic acid | Not found | |
| phenylalanine | 400-4000§ | (Mopper and Lindroth 1982) |
| pyridoxine | 50-450* | (Sañudo-Wilhelmy et al. 2012) |
| riboflavin | 0.2-5* | (Sañudo-Wilhelmy et al. 2012) |
| <i>S</i> -(5'-adenosyl)-L-homocysteine | Not found | |
| taurocholic acid | Not found | |
| thiamin | 25-350* | (Sañudo-Wilhelmy et al. 2012) |
| thymidine | Not found | |
| tryptophan | n.d.-1000§ | (Mopper and Lindroth 1982) |

*-Coastal California (1-800 m)

§-Baltic Sea (1-170 m)

∞ -Connecticut pond (1-17 m)

¶-Northeast Atlantic Ocean (1-4000 m)

•-North Pacific Ocean (5-150 m)

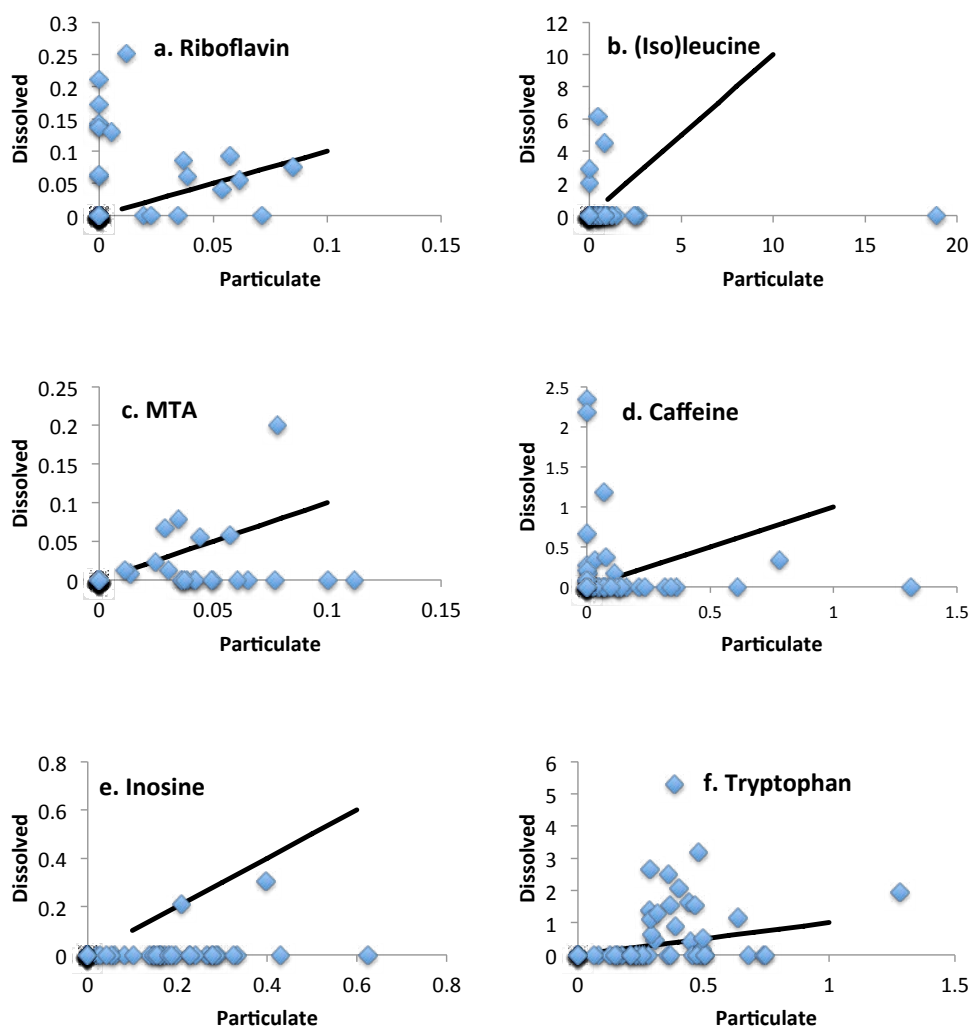


Figure S1. To determine whether there had been significant leakage of metabolites from cells during filtration the particulate ratio of metabolite x / phenylalanine was plotted against the dissolved ratio of metabolite x / phenylalanine. This was done because phenylalanine was measured in most dissolved and particulate samples. Thus, if metabolites were leaking, the ratio in the particulate should equal the ratio in the dissolved sample. The black line shows an equal relationship. This is has been plotted for a) riboflavin, b) (iso)leucine, c) MTA, d) caffeine, e) inosine, f) tryptophan.

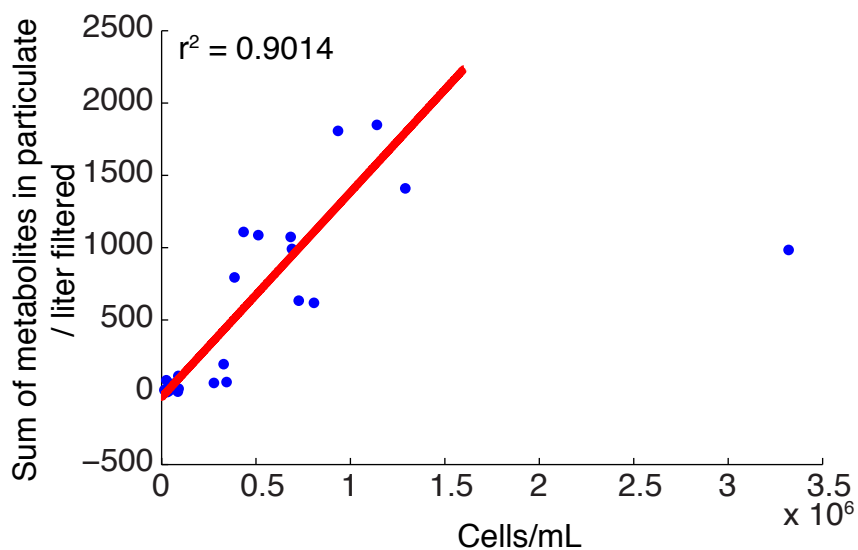


Figure S2. Cell counts plotted versus the total moles of targeted metabolites measured per liter of water filtered. Cell counts were only performed on some samples but these show a linear correlation with the total moles of targeted metabolites measured in the particulate samples suggesting that this is a reasonable way to correct for biomass. One outlier has been discarded.

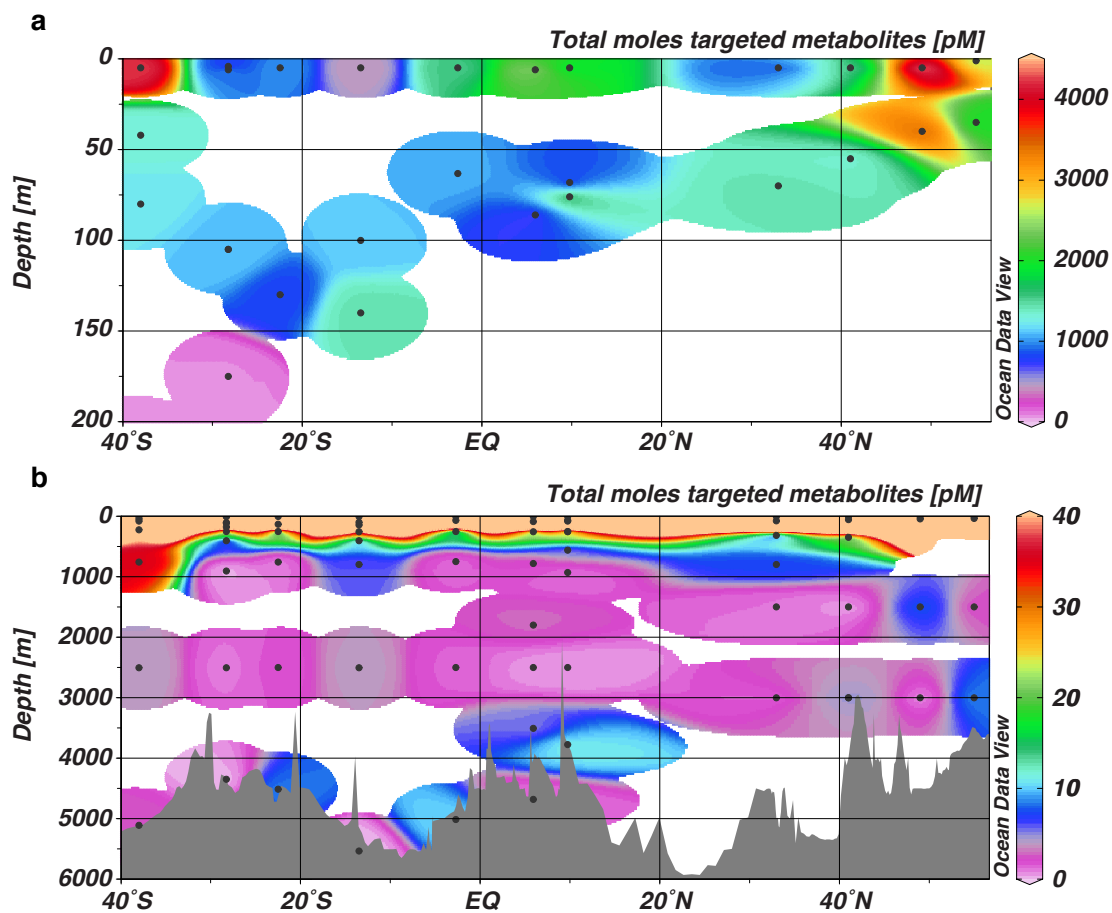


Figure S3. Distribution of total moles of targeted metabolites over the transect, in the top 200 m (a) and at all depths (b).

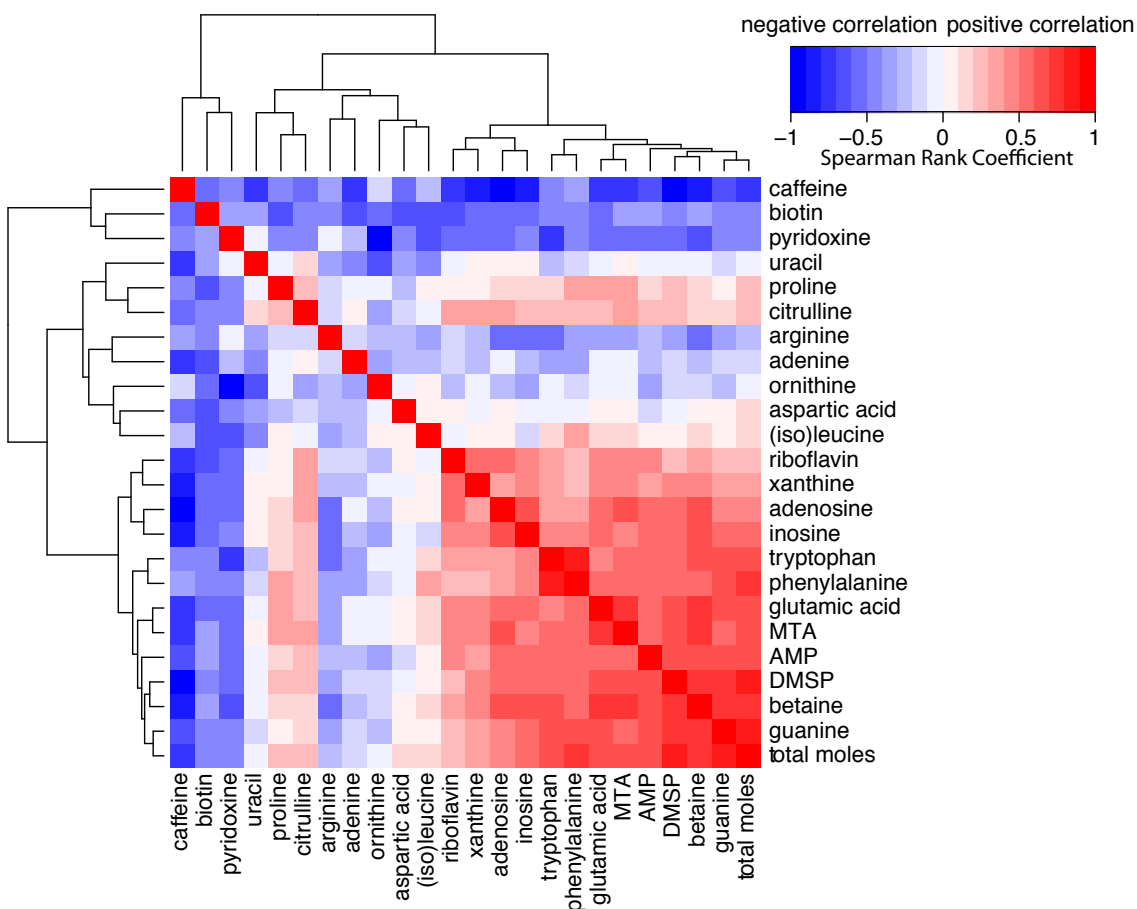


Figure S4. Spearman rank correlation coefficients plotted as a heatmap with hierarchical clustering shows which particulate metabolite distributions co-vary. Red indicates a positive correlation and blue indicates a negative correlation. Total moles of targeted metabolites to show metabolites that are positively and negatively correlated with this proxy for biomass.

Supplemental References

- Amin, S. A., L. R. Hmelo, H. M. van Tol, and others. 2015. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **522**: 98–101. doi:10.1038/nature14488
- Ducklow, H. 2000. Bacterial production and biomass in the oceans, p. 85–120. *In* D.L. Kirchman [ed.], *Microbial Ecology of the Oceans*. Wiley.
- Le Gall, A. C., and C. M. G. Van Den Berg. 1998. Folic acid and glutathione in the water column of the North East Atlantic. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **45**: 1903–1918. doi:10.1016/S0967-0637(98)00042-9
- Hu, H., S. E. Mylon, and G. Benoit. 2006. Distribution of the thiols glutathione and 3-mercaptopropionic acid in Connecticut lakes. *Limnol. Oceanogr.* **51**: 2763–2774. doi:10.4319/lo.2006.51.6.2763
- Mopper, K., and P. Lindroth. 1982. Diel and depth variations in dissolved free amino acids and ammonium in the Baltic Sea determined by shipboard HPLC analysis. *Limnol. Oceanogr.* **27**: 336–347. doi:10.4319/lo.1982.27.2.0336
- Sañudo-Wilhelmy, S. A., L. S. Cutter, R. Durazo, and others. 2012. Multiple B-vitamin depletion in large areas of the coastal ocean. *Proc. Natl. Acad. Sci. U. S. A.* **109**: 14041–5. doi:10.1073/pnas.1208755109

Chapter 5.

Metabolite composition of sinking particles reflects a changing microbial community and differential metabolite degradation

Abstract

Marine sinking particles transport carbon from the surface and bury it in deep sea sediments where it can be sequestered on geologic time scales. The combination of the food web in the surface ocean that produces these particles and the particle-associated microbial community that degrades these particles, creates a complex set of variables that control organic matter production, composition, and degradation. In this study we use targeted metabolomics to characterize a suite of small biomolecules, or metabolites, in sinking particles and compare the metabolite composition of the sinking particles to the suspended particles in the euphotic zone from which they are derived. These samples were collected in the South Atlantic subtropical gyre as well as in the equatorial region and the Amazon River Plume. The total flux of targeted metabolites into the particle traps deployed at 150 m did not shift in tandem with higher surface abundances of phytoplankton, which have been linked to greater carbon export in past studies. Metabolites possibly derived from the degradation of nucleic acids and lipids are detected as an increasing mole fraction of the targeted metabolites in the sinking particles relative to surface suspended particles, while algal-derived metabolites are a smaller fraction of the observed metabolites on the sinking particles.

5.1 Introduction

Sinking particles in the ocean are composed of a complex and varying mixture of dead or dying phytoplankton, zooplankton fecal pellets, aggregates held together by exopolysaccharides, and the microbes degrading this organic matter. These particles are of great interest because they are responsible for delivering organic and inorganic carbon to deep ocean sediments where the carbon can be buried on geologic time scales. Estimates from satellite data and models place the total global flux of carbon out of the euphotic zone at around 6 Pg C yr⁻¹ (Siegel et al. 2014); but the variables that control the magnitude of this flux in different ocean regions are still not fully understood. An important variable is the source and composition of the organic matter associated with these particles and the mechanisms by which it is transformed and remineralized by the particle-associated microbial community (reviewed by Boyd and Trull 2007). Here, we use targeted metabolomics to compare the metabolite composition of sinking particles collected at 150 m to that of the suspended particles (i.e., cells) in the euphotic zone that are the presumed source of these sinking particles.

A complex set of food web interactions, that can vary across oceanic regions, determines the flux of particulate organic matter into the deep ocean (Boyd and Newton 1999). Specifically, characteristics like particle sinking rates appear to play a large role in how much organic matter leaves the euphotic zone. Bacterial respiration of organic matter appears to be inversely related to sinking rates in regions such as the Southern Ocean where rapid sinking rates occur with low respiration; in contrast oligotrophic regions like the North Atlantic subtropical gyre show the opposite trend (McDonnell et al. 2015). Characterization of particulate organic matter at the level of molecular classes has shown that the amino acid fraction is approximately 40-50% of the organic matter, while only 20-30% is carbohydrate and lipid. On a bulk level the relative

contributions of these molecules does not change significantly with depth while the total amount of organic carbon decreases (Hedges et al. 2001).

The consistency of these bulk trends disguises the interactions between molecular-level organic matter composition and its processing by the microbial community that may contribute to variability in carbon export. For instance, while rapidly sinking phytoplankton cells during blooms are generally considered to contribute to a large export flux of organic carbon, fecal pellets produced by grazing can contribute similar or greater total fluxes of organic matter out of the euphotic zone (Wiedmann et al. 2014). These different sources may affect the molecular composition of the sinking particles with implications for the metabolic pathways expressed by the microbial community during organic matter degradation.

Within classes of biomolecules, specific molecules are differentially degraded and produced on sinking particles. For instance, lipid composition of sinking particles changes with depth. Certain types of fatty acids, in particular polyunsaturated fatty acids, are degraded rapidly while other lipid types like long-chain alkenones are not degraded and reach the sediments (Wakeham et al. 1997). Similarly, some free amino acids like threonine, arginine, aspartic acid, glycine, and valine are enriched in deeper, more degraded sinking particle samples suggesting that they are either not degraded as rapidly, or are produced as the particles sink (Ingalls et al. 2003). In contrast, relative monosaccharide composition on sinking particles is similar to the source composition with xylose and fucose decreasing slightly and arabinose increasing as a mole percent of carbohydrates (Tanoue and Handa 1980). While these studies come from many ocean regions and depths, they collectively show that there is measurable variability in concentrations of specific molecules despite relatively small variations in molecular class distributions across sinking marine POM.

Marine particles are hotspots of microbial activity. Microbes actively colonize these particles to exploit the abundance of bioavailable organic molecules (Kjørboe et al. 2002). The community that colonizes these particles has been documented to consist of copiotrophic bacteria known to be associated with dissolved organic matter enrichment or eukaryotic phytoplankton, specifically members of *Alteromonas*, *Methylophaga*, *Flavobacteriales*, and *Pseudoalteromonas*. These microbes are well-adapted to utilize a diverse array of organic molecules (Fontanez et al. 2015) and thus are an important control on the delivery of organic matter to the deep ocean. As particles exit the euphotic zone, the initial microbial community contains species that colonize the particles at the surface, but community succession is observed as particles sink and their composition changes through degradation (Thiele et al. 2015).

In this study, we use targeted metabolomics to characterize the metabolite composition of sinking particles at 150 m in the Western South Atlantic Ocean. Metabolomics simultaneously characterizes a broad suite of small, structurally diverse organic molecules produced by biosynthetic pathways. These molecules include amino acids, nucleic acids, vitamins and cofactors, and biosynthetic intermediates but does not include biopolymers such as proteins, polysaccharides, or DNA/RNA. By comparing the metabolite composition of these sinking particles to the composition of their source particles, we can identify metabolites likely being degraded or produced by the microbial community on the particles. These metabolites also reflect changes in community composition, microbial physiology, and microbial interactions through signaling molecules.

The metabolite fluxes into the particle traps measured here can be contextualized by comparison to the POC flux and the composition (fecal pellets, aggregates, phytoplankton species) of the same sinking particle samples that have been published by Durkin et al. (2016).

Durkin et al. (2016) found that there was not significant variability in the POC flux along this transect, but that there were differences in the particle composition at the stations along the transect. In particular, the authors compared the abundance of different phytoplankton cells in the surface to the cells in the particle trap to determine the fraction of whole cells that reached the trap. In addition, observations of particle trap material yielded an assessment of the dominant particle source according to station (diatom, coccolithophore, or grazing & aggregation influenced). These parameters assist in linking metabolite fluxes into the particle traps to surface sources or degradation products.

5.2 Materials and methods

5.2.1 Field Sampling

Samples were collected during cruise KN210-04 in the western South Atlantic from 25 March – 9 May 2013, transiting from Montevideo, Uruguay to Bridgetown, Barbados. Samples were collected at depths including 5 m (referred to as the surface), the deep chlorophyll maximum (DCM; determined by fluorescence), and 250 m. Sinking particles were collected in a conical net trap that was 2 m in diameter with a cod end (Peterson et al. 2005). It was deployed at 150 m and tethered to the surface. After approximately 24 h, the trap was acoustically closed and retrieved (see Durkin et al. (2016) and Table S1 for deployment times). The sample was split into eight equal parts using a sample splitter (Lamborg et al. 2008) and one split was used for the metabolomics analysis.

5.2.2 Materials

All metabolite standards were purchased from Sigma-Aldrich at the highest purity available with the following exceptions: dimethylsulfoniopropionate (DMSP), which was

purchased from Research Plus, Inc.; 2,3-dihydroxypropane-1-sulfonate (DHPS) and acetyltaurine, which were donated by Dr. Mary Ann Moran (University of Georgia); and *S*-(1,2-dicarboxyethyl)glutathione, which was purchased from Bachem. Hydrochloric acid (trace metal grade), acetonitrile (Optima grade), and methanol (Optima grade), were obtained from Fisher Scientific and formic acid from Fluka Analytical. Glutamic acid- d_3 was bought from Cambridge Isotopes, 4-hydroxybenzoic acid- d_4 from CDN Isotopes, and sodium taurocholate- d_5 from Toronto Research Chemicals through Fisher Scientific. All water was purified by a Milli-Q system (Millipore; resistivity $18.2 \text{ M}\Omega\cdot\text{cm}$ @ 25°C , $\text{TOC} < 1 \text{ }\mu\text{M}$). Glassware used to collect samples and GF/F filters were combusted in an oven at 460°C for at least 5 h before being brought on the cruise and all plasticware was autoclaved before use. The filter holders and tubing for shipboard filtration were flushed with 10% HCl and then Milli-Q water between each sample.

5.2.3 Shipboard sample processing

Water (4 L), for particulate sampling, was collected directly from Niskin bottles into polytetrafluoroethylene (PTFE) or polycarbonate bottles. The water was then filtered through a $0.7 \text{ }\mu\text{m}$ GF/F filter (Whatman) using a peristaltic pump. The split from the particle trap was filtered in the same way, but either through a $1.6 \text{ }\mu\text{m}$ GF/A filter (stations 5, 7, 9, 11, 12, 18) or through a $0.7 \text{ }\mu\text{m}$ GF/F filter (stations 16, 19, 20, 21, 22, 23) due to a shortage of filters. The filters were returned to their combusted aluminum foil wrappers for storage. Filters were stored at -80°C until extraction in the laboratory.

5.2.4 Laboratory sample processing

Filters were extracted as close to the time of analysis on the mass spectrometer as possible ($< 48 \text{ h}$). The filter extraction protocol was adapted from Rabinowitz and Kimball (2007). One half of each filter ($1/4$ of the filter from the net traps) was weighed and cut into

small pieces using methanol-rinsed scissors and tweezers on combusted aluminum foil. The pieces of filter were placed in an 8 mL glass amber vial with 1 mL of -20°C extraction solvent (40:40:20 acetonitrile:methanol:water + 0.1 M formic acid). 25 µL of a 1 µg mL⁻¹ deuterated standard mix (glutamic acid-d₃, 4-hydroxybenzoic acid-d₄, taurocholate-d₅) were added as an extraction recovery standard. Each vial was gently vortexed to separate filter pieces and expose them to the solvent. The vials were then sonicated for 10 min. The extract was transferred with a Pasteur pipette to 1.5 mL Eppendorf centrifuge tubes. The filter pieces that remained in the 8 mL vials were rinsed with 200 µL of cold extraction solvent that was then also added to a centrifuge tube. The extracts were centrifuged at 20,000 x g for 5 min to remove cellular detritus and filter particles. The supernatant was transferred to clean 8 mL glass amber vials. Extracts were neutralized with 26 µL of 6 M ammonium hydroxide. Samples were dried to near complete dryness in a vacufuge. They were then reconstituted with either 247.5 µL 95:5 water:acetonitrile and 2.5 µL biotin-d₂ (injection standard) for samples collected at the surface or DCM or 123.8 µL 95:5 water:acetonitrile and 1.25 µL biotin-d₂ (injection standard) for deeper samples with presumed lower biomass. The net trap samples were reconstituted in 495 µL 95:5 water:acetonitrile and 5 µL biotin-d₂ (injection standard). 100 µL of each solution were loaded into a glass insert in an autosampler vial. 15 µL of each sample type were combined to create a pooled sample for water column samples and a pooled sample for net trap samples.

5.2.5 Mass spectrometry

LC-MS/MS analysis was performed on a Phenomenex C18 column (Synergi Fusion, 2.1 x 150 mm, 4 µm) coupled via heated electrospray ionization (ESI) in polarity switching mode to a triple quadrupole mass spectrometer (Thermo Scientific TSQ Vantage) operated under selected reaction monitoring mode (SRM), as described previously (Kido Soule et al. 2015).

Quantification and confirmation SRM transitions were monitored for each analyte. Eluent A (Milli-Q water with 0.1% formic acid) and Eluent B (acetonitrile with 0.1% formic acid) were run at 250 μ L / min in the following gradient: hold at 5% B for 2 min; ramp to 65% B for 16 min; ramp to 100% B for 7 minutes and hold for 8 min. The column was re-equilibrated with the starting ratio of eluents for 8.5 min between sample analyses. Samples were run in batches of approximately 50 with a pooled sample that corresponded to that batch. Before each batch of samples was run, the column was conditioned with 5 injections of the pooled samples. A pooled QC sample was then run after every 10 samples.

5.2.6 Data processing

XCalibur RAW files generated by the mass spectrometer were converted to mzML files using MSConvert (Chambers et al. 2012). MAVEN (Melamud et al. 2010; Clasquin et al. 2012) was used to select and integrate peaks. Calibration peaks below a MAVEN quality threshold of 0.4 (on a scale of 0-1) and sample peaks below 0.2 were discarded. Quality scores are generated by an machine-learning algorithm that has been trained to recognize the quality of peaks by incorporating information about peaks scored by an expert human analyst (Melamud et al. 2010). To enhance confidence in metabolite identification, quantification and confirmation peaks were required to have retention times within 12 sec (0.2 min) of each other (peak width can range from 10-30 seconds and confirmation ions, which have lower intensity, can have a retention time assigned at a slightly different point on the peak resulting in retention times that can differ by ~0.1 seconds). Confirmation ions were required to have a MAVEN quality score of at least 0.1 and a signal-to-noise ratio greater than 1. Calibration curves were required to have at least five calibration points with the highest calibration point one concentration level above the highest concentration in a sample.

5.2.7 Temperature, salinity, and chlorophyll *a*

The R/V *Knorr* was equipped with a SBE9+ CTD with a depth limit of 6000m. A SBE3T/SBE4C sensor system was used to measure temperature and conductivity. A Wet Labs FLNTURTD combination fluorometer and turbidity sensor was used to detect fluorescence. Fluorescence data was calibrated to direct chlorophyll *a* measurements. Chlorophyll *a* samples were measured by Gwenn Hennon according to methods from Arar & Collins (1997).

5.2.8 Calculations and computational tools

Matlab (R2014a; MathWorks, Natick, MA) was used to process data and create figures. The map and supplemental profile images were created with Ocean Data View (Schlitzer 2016). The non-metric multidimensional scaling joint plot was calculated and plotted using the R stats package (R Core Team 2015). Metabolite fluxes into the particle traps were calculated using the following:

$$F = \frac{[M]_t}{A \times t_d}$$

where F is the flux ($\text{nmol m}^{-2} \text{d}^{-1}$), $[M]_t$ is the abundance of a metabolite M captured in the net trap (nmol), A is the area of trap opening (3.14 m^2), and t_d is the trap deployment time (days). The export turnover rates (TR) of metabolites that were measured at both the surface and in the net trap at a given station were calculated using:

$$TR = \frac{F}{[M]_s \times Z_{ML}}$$

where TR is the export turnover rate (d^{-1}), $[M]_s$ is the concentration of metabolite M in the surface (nmol m^{-3}), and Z_{ML} is the mixed layer depth (m). Z_{ML} was calculated by Durkin et al. (2016) as the depth at which down-cast temperatures increased at a rate greater than $0.15^\circ\text{C m}^{-1}$ (see Table S1 for mixed layer depth at each station).

5.3 Results

5.3.1 Sampling

Samples were collected along a latitudinal transect in the Western South Atlantic Ocean. This encompassed ocean regions including the South Atlantic Subtropical Gyre, the equatorial region, and one station in the Amazon River Plume (Fig. 1). These regions could be distinguished on the basis of the salinity profiles (Fig. S1). There were five stations where targeted metabolites were measured in both the suspended particles at the surface, DCM, and 250 m as well as in the sinking particles (see Fig. S2 for chlorophyll *a* profile). In addition, there were seven stations where only the metabolites in the sinking particles were measured and two stations with exclusively water column data. While suspended and sinking particles could not be compared directly at all stations, overall compositional differences between suspended and sinking particles could be examined. The particle trap samples contained greater amounts of material than the suspended particle samples (Fig. S3) and so comparisons between them are based on relative targeted metabolite composition, as calculated by normalizing moles of each metabolite to the total moles of targeted metabolites measured. As noted in the methods, all suspended particles were captured on a 0.7 μm filter, but about half the sinking particle samples were captured on a 1.6 μm filter and the other half on a 0.7 μm filter. This difference does not appear to cause a systematic effect on the metabolite abundances (Fig. 2), perhaps because the quantity and size of the particles immediately blocked the filter and the nominal pore size was likely less than 0.7 μm on both filter types.

5.3.2 Metabolite composition of sinking particles

72 metabolites passed all quality controls (see data processing methods), but only 42 of

these metabolites were detected (see Chap. 2 for detection limits) in the sinking particle samples (Table S2). There were 18 metabolites that had a flux into the particle traps greater than 1 nmol m⁻² d⁻¹ at one or more stations (Fig. 2). These metabolites included osmolytes, amino acids, and a nucleotide. The total flux of these 18 metabolites into the particle traps at 150 m shows some variability along the transect with station 23 having the lowest flux and station 16 having the highest flux. Betaine had the largest flux of any metabolite at all stations while other metabolite fluxes were more variable. For example, guanine is a larger component of the flux into the net trap at station 9 while DMSP and proline contribute more to the flux at station 16. Choline is also a sizable component of the flux at many stations, in a number of cases on par with DMSP. Other metabolites such as 5-methylthioadenosine (MTA), adenosine, caffeine, indole 3-acetic acid (IAA), inosine, (iso)leucine, phenylalanine, riboflavin, tryptophan, and xanthine contribute much smaller total moles to the flux of these metabolites but are measured throughout the transect (Table 1). While the total contribution of these metabolites to the POC flux is quite low, with the highest contribution being ~0.07% by weight (Fig. S4), they are indicative of the processes occurring on the sinking particles. Durkin et al. (2016) designated certain stations as dominated by diatoms (stns 12, 23), coccolithophores (stns 9, 21), or aggregates (mixtures of detrital material; stns 7, 16, 19). Our data on the overall fluxes, as well as the fluxes of individual metabolites, did not show a relationship to these groups.

5.3.3 Comparison of suspended and sinking particles

Most metabolites that were measured in particles filtered from the water column were also measured in the particle trap samples (Table 1). Considering that the particle trap samples contained a great deal more material than the water column samples, it was surprising that there was only one metabolite that was not detected in any water column samples but was detected in

many particle trap samples. This metabolite was indole 3-acetic acid (IAA), a phytoplankton growth hormone. In contrast, cysteine, D-ribose 5-phosphate, ornithine, and orotic acid were all measured in at least one water column sample but were not detected in any particle trap samples.

When the targeted metabolite composition (metabolite concentrations normalized to the total concentration of targeted metabolites) of suspended water column particles is compared to that of sinking particles using non-metric multidimensional scaling, three clusters emerge (Fig. 3). Suspended particles from the surface and DCM form a group that is continuous with the suspended particles collected at 250 m, while the sinking particles cluster separately. The only exception is the DCM sample from station 15, which clusters more closely to the sinking particle samples. This clustering is driven by the relative amounts of certain metabolites in these samples. For instance, DMSP is a larger mole fraction of suspended particles in the surface and DCM, while phenylalanine and, in particular, guanine make up more of the metabolite pool in deeper suspended particle samples around 250 m. There are a variety of metabolites that are a greater share of the metabolite pool associated with sinking particles compared to the suspended particles. Most notably, these include betaine, adenosine, xanthine, and indole 3-acetic acid (Fig. 3).

A greater number of metabolites have larger mole fractions in the sinking particles in part due to the fact that DMSP is such a large portion of the total moles measured in surface and DCM samples (Fig. 4b), with DMSP reaching upwards of 80% of the total moles of targeted metabolites measured. In contrast, while still present in the particle traps, DMSP is only ~20-30% of the total targeted moles measured. In an almost inverse pattern, betaine reaches between 50-70% of the total composition of the sinking particles while being a much smaller component of the suspended water column samples (Fig. 4a). Proline, an amino acid that is also a higher

mole fraction on the sinking particles than in suspended particles (Fig. 4c), also persists in deep (thousands of meters) suspended particle samples as a relatively large mole fraction (as shown in Chapter 4). Finally, a nucleotide derivative, xanthine, while a small portion of the total metabolite pool, is nonetheless elevated on sinking particles relative to suspended particles (Fig. 4d). However, its mole fraction is much more variable across the transect than the more abundant metabolites.

5.3.4 Metabolite export turnover rates

To estimate the fraction of surface-derived metabolites that were delivered to 150 m via sinking particles, export turnover rates were calculated for individual metabolites. This calculation determines whether there is differential export of certain metabolites from the euphotic zone. To make this calculation we must assume that the metabolite is produced in the surface and subsequently removed as the particles sink, which is not necessarily the case. In this context, the export turnover rate can be interpreted as a relative spectrum with high values indicating more production relative to degradation on the sinking particles and low values having greater amounts of degradation than production. It must be recognized that for most metabolites both processes are likely occurring on sinking particles.

The export turnover rate was calculated for each metabolite that was measured in both the surface and particle trap samples at the five stations where water column and particle trap data were both available (stations 5, 7, 9, 19, and 23). The turnover rate of a metabolite could only be calculated if it was detected at the surface and in the net trap, thus a turnover rate is not reported for every metabolite at each station. The export turnover rates of metabolites ranged from 1×10^{-5} – $2 \times 10^{-1} \text{ d}^{-1}$ (geometric mean: $1.4 \times 10^{-3} \text{ d}^{-1}$) (Table S3; Fig. 5). In comparison, phytoplankton cells had export turnover rates that were generally slightly lower ranging from 2×10^{-6} – 3×10^{-3}

d^{-1} (geometric mean: $8 \times 10^{-5} d^{-1}$) (Durkin et al. 2016). These values varied by an order of magnitude at different stations for some metabolites. These values can potentially indicate which metabolites are predominantly being degraded (low values) versus those whose abundances are most likely controlled by production by the microbial community on the particles as they sink (high values).

For example, 2,3-dihydroxypropane-1-sulfonate (DHPS) had the lowest turnover rates of any metabolite in stations 9 and 19 where its export turnover rate was calculated. This suggests that a lower fraction of surface-produced DHPS reaches 150 m through transport on these sinking particles compared to other metabolites. In contrast, the metabolites with the highest export turnover rates vary by station but include proline, guanine, inosine, xanthine, and betaine at different stations across the transect (Fig. 5). Because metabolites on sinking particles might also be derived from the DCM (all particle traps (150 m) were below the DCM), the same export turnover rate calculation was performed using the metabolite abundances from the DCM (Fig. S5). For the most part, the trends remain the same with, for instance, xanthine and betaine having higher export turnover rates and DMSP and DHPS having low export turnover rates. The one noticeable exception is guanine, which has a lower export turnover rate when calculated at the DCM. This is expected because guanine concentrations are higher at the DCM than at the surface (Chapter 4).

For stations 7, 9, 19, and 23, Durkin et al. (2016) calculated the export turnover rate (based on surface abundance) for different phytoplankton cell types including diatoms, coccolithophores, dinoflagellates, and nanocells (referring to all photosynthetic cells that can be observed with 400x magnification under a microscope). These export turnover rates can be compared to those calculated above for the metabolites (Fig. 5). The phytoplankton export

turnover rates provide a baseline comparison as their source is constrained to the euphotic zone and they are not produced at depth, while the metabolite export turnover rates reflect a spectrum of production versus degradation processes on the sinking particles. The export turnover rates of the phytoplankton are in the middle to low end of the range calculated for metabolites. In particular, DMSP seems to align closely with the phytoplankton turnover rates. In contrast, the higher export turnover rates of the metabolites mentioned above might be the result of production by particle-associated microbes during transport to 150 m. DHPS aligns with the dinoflagellate turnover rate in station 9 and is below any phytoplankton turnover rates in station 23. Station 23 in general has a lower and narrower range of metabolite turnover rates than the other stations suggesting that a smaller fraction of the surface-derived metabolites reach 150 m. This is observed in the export turnover rates of the phytoplankton cells in this station as well. The lower total flux of targeted metabolites at station 23 also supports the observation that a smaller fraction of metabolites reach 150 m (Fig. 2). The wide range of export turnover rates calculated for the different metabolites at any given station suggests that certain metabolites are degraded while others are produced as particles sink from the surface ocean.

5.4 Discussion

5.4.1 Variability in total metabolite flux to 150 m

The total fluxes of targeted metabolites into these particle net traps placed at 150 m are variable along the transect (Fig. 2), despite the fact that total POC fluxes were not significantly different (Durkin et al. 2016). As an example, station 9 and station 23 are located in distinct regions of the transect with station 9 in the oligotrophic gyre and station 23 in the Amazon River plume, and their POC fluxes are 38 ± 5 and $43 \text{ mg m}^{-2} \text{ d}^{-1}$, respectively. However, the total flux

of targeted metabolites into the particle traps at station 9 is $\sim 160 \text{ nmoles m}^{-2} \text{ d}^{-1}$ but at station 23 it is only $\sim 40 \text{ nmoles m}^{-2} \text{ d}^{-1}$ (Fig. 2). This is the inverse of what might be expected, as station 9 is oligotrophic and had less phytoplankton source material compared to station 23 where there was a diatom bloom. Perhaps the most likely explanation is that many of the metabolites measured on the sinking particles at station 9 are derived, not from source material in the euphotic zone, but from the metabolism of the particle-associated microbial community as the particles sink. At station 23 the reverse may be true and the metabolites that reach 150 m could be derived from the sinking phytoplankton. The similarity in POC flux may be explained by a greater input of diatom-derived biopolymers that are not detected in this method (Hoagland et al. 1993). This is supported by the fact that Station 9 has fewer metabolites with calculated export turnover rates (Fig. 5; Table S3) due to the absence of detectable amounts of many metabolites in the surface water, suggesting that the greater flux of metabolites into the particle traps cannot be derived from the surface. In contrast, at station 23 more metabolites are measured in the surface and the export turnover rates were mostly quite close to those of the phytoplankton cells, in particular the diatoms (Fig. 5). This suggests that metabolite delivery to 150 m may have a larger contribution from sinking phytoplankton cells at station 23, while metabolites at station 9 may be derived in greater part from particle-associated microbes. This could be explored further with more information about the microbial community composition of these particles as well as the production and degradation rates of the metabolites. Alternatively, if the sinking rate of the particles was low, then the surface conditions may be decoupled from the captured sinking particles captured, thus yielding lower export turnover rates.

5.4.2 Fluctuations of osmolytes on sinking particles

Compositionally, betaine, DMSP, choline, and guanine contribute the largest share of the

total moles of targeted metabolites measured on these sinking particles (Fig. 2). Betaine and DMSP are both osmolytes, which are molecules present in the cytosol of cells that regulate osmotic pressure as well as assist in maintaining protein structure (Yancey et al. 1982).. Betaine, by far the largest contributor to the metabolite composition of the sinking particles, is an osmolyte utilized by organisms across the domains of life, although it is less commonly used in phytoplankton than DMSP (Yancey et al. 1982; Boch et al. 1997; Keller et al. 1999; Lai and Lai 2011). Betaine concentrations can reach 170 mM inside a phytoplankton cell (Keller et al. 1999), suggesting that this metabolite is likely to be measured at relatively high concentrations in particulate samples. However, in the surface ocean and the DCM, DMSP, an exclusively algal metabolite, has the highest concentrations (Fig. 4b). This trend changes on sinking particles where DMSP is a smaller component than betaine (Fig. 4a), although it is measured in every sample. This is likely due to the degradation of DMSP by heterotrophic microbes as these particles sink from the surface. Heterotrophic bacteria can catabolize DMSP (Reisch et al. 2011) as well as incorporate the carbon and sulfur into biomass (Simó et al. 2009). In addition, DMSP may act as a chemoattractant to guide bacteria to these particles (Seymour et al. 2010).

We hypothesize that this compositional shift from DMSP to betaine is due both to the degradation of DMSP as particles sink, as well as to the increased prevalence of heterotrophic microbes that synthesize betaine, rather than DMSP. The export turnover rates lend further support to this idea, indicating that DMSP has similar export turnover rates to those of phytoplankton cells while, at stations where it is measured in the surface at all, betaine has a higher export turnover rate than the phytoplankton cells (Fig. 5). Having a higher export turnover rate than the cells that might transport these metabolites to depth, suggests that betaine is being generated on the particles themselves as they sink. This indicates both that DMSP availability

might decrease with depth at a similar rate to that of whole phytoplankton cells and that betaine becomes a more abundant metabolite with depth. At 250 m, however, betaine is not detected on suspended particles (Fig. 4a), suggesting that its production may be specific to sinking particles. One possible scenario is that choline, derived from the breakdown of phosphatidylcholine, a lipid known to be present in the surface ocean (Van Mooy and Fredricks 2010) could contribute to the utilization of betaine as a dominant osmolyte on these particles. Some bacteria are known to require exogenous choline as a precursor for the synthesis of betaine (Boch et al. 1994). Another possibility is that direct degradation of betaine lipids, also observed in the surface ocean (Van Mooy and Fredricks 2010), might contribute to the pool of betaine observed on these sinking particles. Either of these scenarios would suggest that lipid degradation on these particles might influence microbial adaptive strategies. Alternatively, betaine might simply be a more frequently utilized osmolyte in the microbial community associated with sinking particles, perhaps not only due to the species composition but also due to greater nitrogen availability on sinking particles. A comparison of the removal of phosphatidylcholine and betaine lipids with the emergence of choline and betaine in the upper 150 m of the water column could test this hypothesis. In addition, analysis of transcripts or enzymes associated with lipid degradation and betaine synthesis might facilitate determining a link.

5.4.3 Accumulation of nucleic acid waste products on sinking particles

Other metabolites that were a higher mole fraction of the targeted metabolites in the sinking particles relative to the water column particles included xanthine, uracil, and caffeine (Fig. 3). Xanthine and caffeine are both purines, and structurally related to the nucleobases guanine and adenine. Xanthine in particular is part of a pathway that recycles components of nucleic acids and can form as a degradation product of RNA or DNA. Given the prevalence of

guanine in suspended particle samples, this might be a source of xanthine. Uracil is a pyrimidine nucleobase that is incorporated into RNA and is structurally related to thymine and cytosine.

Uracil and xanthine both accumulate extracellularly in *Escherichia coli* cultures when they reach stationary phase. This is attributed, in particular, to the degradation of ribosomal RNA (Rinas et al. 1995). The fact that these metabolites accumulate when a bacterial culture reaches a point when nutrients begin to limit growth and dying cells equal dividing cells suggests that these molecules could be linked to the degradation of cellular material. Xanthine, in particular, is a larger mole fraction of the sinking particles than the suspended particles (Fig. 4d) and also is measured in all of the sinking particle samples but not all of the suspended particles from the surface and DCM. At station 19, where a xanthine export turnover rate is calculated, it is higher than any of the other metabolites or phytoplankton cells suggesting that a large fraction of xanthine from the surface reaches 150 m or that it is produced as these particles sink.

Additionally, the xanthine export turnover rate calculated with the DCM metabolite abundance was the highest of any metabolite in the stations where it was calculated (Fig. S5). Not only is xanthine a degradation product, it is also a substrate for the enzyme xanthine oxidase which oxidizes xanthine uric acid and as a result of this reaction produces the reactive oxygen species H_2O_2 and O_2^- (Kelley et al. 2010). In terrestrial environments these reactive oxygen species are used by organisms like fungi to degrade organic substrates like lignin (Dashtban et al. 2010). The gene for this enzyme is present in metagenome studies on sinking particles (Fontanez et al. 2015). In a sinking particle environment, xanthine might be used as a substrate to produce reactive oxygen species to aid in accessing high molecular weight organic matter. Ultimately, as a nitrogen-containing intermediate in purine biosynthesis, it seems unlikely that this metabolite would be preserved on particles. Rather its high concentrations are likely generated on these

sinking particles, either as a product of degradation but also perhaps as an intermediate in further organic matter degradation.

5.4.4 Amino acid dynamics on sinking particles and degradation indices

Amino acids are a molecule class that has been incorporated into organic matter degradation indices by previous researchers. Certain amino acids seem to persist with increased organic matter degradation while others are preferentially removed (Dauwe et al. 1999). This index is based on total hydrolysable amino acids (both free and protein-associated amino acids). The amino acids measured in this study only include free amino acids. For example, arginine, which is an amino acid correlated with increasing organic matter degradation is not detected in the majority of the particle trap samples (Table 1). Arginine is, however, measured in suspended particle samples collected thousands of meters deep in the ocean (Chapter 4). Free amino acids that are prevalent in the sinking particles include (iso)leucine, phenylalanine, proline, and tryptophan (Fig. 2). Phenylalanine and (iso)leucine are both anti-correlated with increased organic matter degradation, while proline and tryptophan are not included in the index (Dauwe et al. 1999). Thus, the free amino acids do not seem to have the same pattern of preservation as the total hydrolysable amino acids. This could be indicative of degradation of proteins that release these individual amino acids or simply that these are common free amino acids as they are also measured throughout the water column in the suspended particles.

Proline is particularly intriguing as its abundance varies substantially across the particle net trap samples as well as having a larger flux than any other amino acid (Fig. 2). At station 16, proline is a larger component of the flux than guanine or choline, contributing around 10% of the total moles. Proline is measured in deeper ocean suspended particle samples, as discussed in Chapter 4. Few measurements of proline have been made in the ocean and recently high amounts

of proline relative to other amino acids were found on bacteria-sized particles at 1000 m (Takasu and Nagata 2015). Additionally, proline's export turnover rate at a number of stations is among the higher export turnover rates measured for the metabolites (Fig. 5). This suggests increased production of this amino acid by microbes at depth both on sinking particles as well as suspended particles.

5.4.5 Metabolites with high and low export turnover rates

Guanine was also a relatively large component of the metabolite composition of the sinking particles. Guanine, a purine nucleobase (building block of DNA), is ubiquitous throughout the suspended particles in the water column. It becomes a larger mole fraction of samples with increasing depth (Fig. 3; Chapter 4). In Chapter 4, its prevalence throughout the ocean was attributed to its essential role in DNA. Interestingly, guanine flux into the net traps is quite variable compared to a metabolite like betaine that seems to be a relatively constant mole fraction of the targeted metabolite flux. For instance, at station 9, guanine has the second largest flux after betaine, but at station 16 DMSP, proline, and betaine all have larger fluxes into the particle traps (Fig. 2). There are surprisingly few reports of the intracellular concentrations of guanine or how much it tends to vary within and among microbes, so it is unclear whether higher amounts of guanine are indicative of greater biomass or of a physiological response. Guanine has relatively high export turnover rates, which supports the idea that it is produced on the sinking particles rather than derived from surface production. In contrast, when the export turnover rate is calculated with the DCM abundance, it has a low export turnover rate due to its much higher abundance in the DCM. Since the source of guanine is unclear, it is difficult to determine which export turnover rate is more representative of the fraction of guanine generated in the euphotic zone that reaches 150 m.

DHPS is an organosulfur metabolite that can be produced by diatoms and metabolized by heterotrophic marine bacteria such as *Ruegeria pomeroyi* (Durham et al. 2015). DHPS is not measured in most suspended particle samples, but it is measured in slightly more than half of the sinking particle samples. Notably, at stations where it was detected in both surface suspended particles and sinking particles, it has a much lower export turnover rate than any other metabolites (Fig. 5) suggesting that a smaller fraction of DHPS reaches 150 m on sinking particles than other metabolites. It is interesting that its removal appears to be even greater than that of DMSP, although DMSP has the second lowest export turnover rate, as they are both organosulfur carbon substrates that can be catabolized by heterotrophs. The fact that these organosulfur molecules seem to be rapidly removed relative to other metabolites supports findings showing that relatively little organic sulfur is delivered to the dissolved pool of organic sulfur below 200 m (Ksionzek et al. 2016).

5.4.6 An auxin on sinking particles

Finally, while indole 3-acetic acid (IAA) is not measured in any suspended particle samples, it is a small but prevalent component of the sinking particle metabolite pool. This may simply be due to the larger organic matter sample sizes captured in the particle traps (Fig. S3). However, it is interesting to note its presence, as it is an auxin responsible for stimulating phytoplankton growth (Evans and Trewavas 1991; Lambrecht et al. 2000). In a phytoplankton-bacterium interaction, a heterotrophic bacterium produces IAA to augment phytoplankton growth. It synthesizes IAA with tryptophan obtained from the phytoplankton (Amin et al. 2015). There is not a linear relationship between IAA and tryptophan on the sinking particles, perhaps due to continued production of IAA despite the lack of viable phytoplankton (Durkin et al. 2016).

5.5 Conclusions

The targeted metabolites measured on sinking particles are representative of only a small portion of the organic matter being exported from the surface, but it is a dynamic and labile component of POC that reflects a combination of the surface source as well as the microbial processing of sinking particulate organic matter. Unexpectedly, the total flux of targeted metabolites into the particle traps was not linked to POC flux, oceanic region, or dominant phytoplankton cell type (Durkin et al. 2016). This suggests that the particle-associated microbial community may play an important role in the dynamics of metabolites on sinking particles. Metabolites that comprise a higher mole fraction of the sinking particles, such as betaine and xanthine, may be indicative of the types of molecules produced by the particle-associated microbial community. They may also indicate changes in the types of organic substrates available to successive microbial communities on the particles as they sink. In contrast, algal metabolites like DMSP and DHPS are degraded more quickly than other metabolites or are simply not regenerated as the algal metabolic signature is removed from the particles.

The metabolite composition of the sinking particles yields insight into the dynamics of metabolites with different sources, including the euphotic zone and the particle-associated microbial communities. It is difficult to distinguish the source of individual metabolites due to the range of production and degradation processes that are shaping their abundances on sinking particles. Some metabolites can be linked to certain sources because they are known to be derived from degradation of specific organic molecules while others are exclusively derived from certain species. Nevertheless, most have complex roles to play in many organisms making their environmental variability challenging to parse. Moving forward we need more samples to

confirm the trends in metabolite export in larger datasets. Linking our dataset to further information on the microbial community composition and the transcription of certain genes, as well as obtaining further data on the production and degradation rates of metabolites, will allow us to link some of these metabolite dynamics to specific microbial metabolic capabilities and pathways.

Acknowledgments

Funding for this work came from a National Science Foundation grant (Grant OCE-1154320 to EBK and KL) and a WHOI Ocean Ventures Fund award to WMJ. The instruments in the WHOI FT-MS Facility were purchased with support from the GBMF and NSF. Support for WMJ was provided by a National Defense Science and Engineering Fellowship. I would like to thank the captain and crew of the R/V *Knorr*, William Arnold, Benjamin Van Mooy, Justin Ossolinski, Colleen Durkin, Melissa Kido Soule, Krista Longnecker, and Catherine Carmichael for helping to make this dataset possible.

References

- Amin, S. A., L. R. Hmelo, H. M. van Tol, and others. 2015. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **522**: 98–101. doi:10.1038/nature14488
- Arar, E. J., and G. B. Collins. 1997. Method 445.0: *In vitro* determination of chlorophyll *a* and pheophytin *a* in marine and freshwater algae by fluorescence.
- Boch, J., B. Kempf, and E. Bremer. 1994. Osmoregulation in *Bacillus subtilis*: Synthesis of the osmoprotectant glycine betaine from exogenously provided choline. *J. Bacteriol.* **176**: 5364–5371.
- Boch, J., G. Nau-Wagner, S. Kneip, and E. Bremer. 1997. Glycine betaine aldehyde dehydrogenase from *Bacillus subtilis*: Characterization of an enzyme required for the synthesis of the osmoprotectant glycine betaine. *Arch. Microbiol.* **168**: 282–289. doi:10.1007/s002030050500
- Boyd, P. W., and P. P. Newton. 1999. Does planktonic community structure determine downward particulate organic carbon flux in different oceanic provinces? *Deep. Res. Part I Oceanogr. Res. Pap.* **46**: 63–91. doi:10.1016/S0967-0637(98)00066-1
- Boyd, P. W., and T. W. Trull. 2007. Understanding the export of biogenic particles in oceanic waters: Is there consensus? *Prog. Oceanogr.* **72**: 276–312. doi:10.1016/j.pocean.2006.10.007
- Chambers, M. C., B. Maclean, R. Burke, and others. 2012. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**: 918–920. doi:10.1038/nbt.2377
- Clasquin, M. F., E. Melamud, and J. D. Rabinowitz. 2012. LC-MS data processing with MAVEN: A metabolomic analysis and visualization engine. *Curr. Protoc. Bioinforma.* **37**: 1–23. doi:10.1002/0471250953.bi1411s37
- Dashtban, M., H. Schraft, T. A. Syed, and W. Qin. 2010. Fungal biodegradation and enzymatic modification of lignin. *Int. J. Biochem. Mol. Biol.* **1**: 36–50.
- Dauwe, B., J. J. Middelburg, P. M. J. Herman, and C. H. R. Heip. 1999. Linking diagenetic alteration of amino acids and bulk organic matter reactivity. *Limnol. Oceanogr.* **44**: 1809–1814. doi:10.4319/lo.1999.44.7.1809
- Durham, B. P., S. Sharma, H. Luo, and others. 2015. Cryptic carbon and sulfur cycling between surface ocean plankton. *Proc. Natl. Acad. Sci. U. S. A.* **112**: 453–457. doi:10.1073/pnas.1413137112
- Durkin, C. A., B. A. S. Van Mooy, S. T. Dyhrman, and K. O. Buesseler. 2016. Sinking phytoplankton associated with carbon flux in the Atlantic Ocean. *Limnol. Oceanogr.* doi:10.1002/lno.10253
- Evans, L., and A. Trewavas. 1991. Is algal development controlled by plant growth substances? *J. Phycol.* **27**: 322–326. doi:10.1111/j.0022-3646.1991.00322.x
- Fontanez, K. M., J. M. Eppley, T. J. Samo, D. M. Karl, and E. F. Delong. 2015. Microbial community structure and function on sinking particles in the North Pacific Subtropical Gyre. *Front. Microbiol.* **6**: 1–14. doi:10.3389/fmicb.2015.00469
- Hedges, J. I., J. a Baldock, Y. G  linas, C. Lee, M. Peterson, and S. G. Wakeham. 2001. Evidence for non-selective preservation of organic matter in sinking marine particles. *Nature* **409**: 801–4. doi:10.1038/35057247
- Hoagland, K. D., J. R. Rosowski, M. R. Gretz, and S. C. Roemer. 1993. Diatom extracellular polymeric substances: Function, fine structure, chemistry, and physiology. *J. Phycol.* **29**:

- 537–566.
- Ingalls, A. E., C. Lee, S. G. Wakeham, and J. I. Hedges. 2003. The role of biominerals in the sinking flux and preservation of amino acids in the Southern Ocean along 170°W. *Deep. Res. Part II Top. Stud. Oceanogr.* **50**: 713–738. doi:10.1016/S0967-0645(02)00592-1
- Keller, M. D., R. P. Kiene, P. A. Matrai, and W. K. Bellows. 1999. Production of glycine betaine and dimethylsulfoniopropionate in marine phytoplankton. I. Batch cultures. *Mar. Biol.* **135**: 237–248. doi:10.1007/s002270050621
- Kelley, E. E., N. K. H. Khoo, N. J. Hundley, U. Z. Malik, B. A. Freeman, and M. M. Tarpey. 2010. Hydrogen peroxide is the major oxidant product of xanthine oxidase. *Free Radic. Biol. Med.* **48**: 493–498. doi:10.1016/j.freeradbiomed.2009.11.012
- Kido Soule, M. C., K. Longnecker, W. M. Johnson, and E. B. Kujawinski. 2015. Environmental metabolomics: Analytical strategies. *Mar. Chem.* **177**: 374–387. doi:10.1016/j.marchem.2015.06.029
- Kjørboe, T., H.-P. Grossart, H. Ploug, and K. Tang. 2002. Mechanisms and rates of bacterial colonization of sinking aggregates. *Appl. Environ. Microbiol.* **68**: 3996–4006. doi:10.1128/AEM.68.8.3996-4006.2002
- Ksionzek, K. B., K. B. Ksionzek, O. J. Lechtenfeld, S. L. Mccallister, P. Schmitt-kopplin, J. K. Geuer, W. Geibert, and B. P. Koch. 2016. Dissolved organic sulfur in the ocean : Biogeochemistry of a petagram inventory. **7796**.
- Lai, S. J., and M. C. Lai. 2011. Characterization and regulation of the osmolyte betaine synthesizing enzymes GSMT and SDMT from halophilic methanogen *Methanohalophilus portucalensis*. *PLoS One* **6**. doi:10.1371/journal.pone.0025090
- Lamborg, C. H., K. O. Buesseler, J. Valdes, and others. 2008. The flux of bio- and lithogenic material associated with sinking particles in the mesopelagic “twilight zone” of the northwest and North Central Pacific Ocean. *Deep. Res. Part II Top. Stud. Oceanogr.* **55**: 1540–1563. doi:10.1016/j.dsr2.2008.04.011
- Lambrecht, M., Y. Okon, A. Vande Broek, and J. Vanderleyden. 2000. Indole-3-acetic acid: A reciprocal signalling molecule in bacteria-plant interactions. *Trends Microbiol.* **8**: 298–300. doi:10.1016/S0966-842X(00)01732-7
- McDonnell, A. M. P., P. W. Boyd, and K. O. Buesseler. 2015. Effects of sinking velocities and microbial respiration rates on the attenuation of particulate carbon fluxes through the mesopelagic zone. *Global Biogeochem. Cycles* **29**: 175–193. doi:10.1002/2014GB004935.Received
- Melamud, E., L. Vastag, and J. D. Rabinowitz. 2010. Metabolomic analysis and visualization engine for LC-MS data. *Anal. Chem.* **82**: 9818–9826.
- Van Mooy, B. A. S., and H. F. Fredricks. 2010. Bacterial and eukaryotic intact polar lipids in the eastern subtropical South Pacific: Water-column distribution, planktonic sources, and fatty acid composition. *Geochim. Cosmochim. Acta* **74**: 6499–6516. doi:10.1016/j.gca.2010.08.026
- Peterson, M. L., S. G. Wakeham, C. Lee, M. a. Askea, and J. C. Miquel. 2005. Novel techniques for collection of sinking particles in the ocean and determining their settling rates. *Limnol. Oceanogr. Methods* **3**: 520–532. doi:10.4319/lom.2005.3.520
- Rabinowitz, J. D., and E. Kimball. 2007. Acidic acetonitrile for cellular metabolome extraction from *Escherichia coli*. *Anal. Chem.* **79**: 6167–6173.
- R Core Team. 2015. R: A language and environment for statistical computing.
- Reisch, C. R., M. A. Moran, and W. B. Whitman. 2011. Bacterial catabolism of

- dimethylsulfoniopropionate (DMSP). *Front. Microbiol.* **2**: 172–184. doi:10.3389/fmicb.2011.00172
- Rinas, U., K. Hellmuth, R. Kang, A. Seeger, and H. Schlieker. 1995. Entry of *Escherichia coli* into stationary phase is indicated by endogenous and exogenous accumulation of nucleobases. *Appl. Environ. Microbiol.* **61**: 4147–4151.
- Schlitzer, R. 2016. Ocean Data View.
- Seymour, J., R. Simó, T. Ahmed, and R. Stocker. 2010. Chemoattraction to dimethylsulfoniopropionate throughout the marine microbial food web. *Science* (80-.). **329**: 342–345.
- Siegel, D. A., K. O. Buesseler, S. C. Doney, S. F. Sailley, M. J. Behrenfeld, and P. W. Boyd. 2014. Global assessment of ocean carbon export by combining satellite observations and food-web models. *Global Biogeochem. Cycles* **28**: 181–196. doi:10.1002/2013GB004743.Received
- Simó, R., M. Vila-Costa, L. Alonso-Sáez, C. Cardelús, Ò. Guadayol, E. Vázquez-Domínguez, and J. Gasol. 2009. Annual DMSP contribution to S and C fluxes through phytoplankton and bacterioplankton in a NW Mediterranean coastal site. *Aquat. Microb. Ecol.* **57**: 43–55. doi:10.3354/ame01325
- Takasu, H., and T. Nagata. 2015. High proline content of bacteria-sized particles in the Western North Pacific and its potential as a new biogeochemical indicator of organic matter diagenesis. *Front. Mar. Sci.* **2**: 110. doi:10.3389/fmars.2015.00110
- Tanoue, E., and N. Handa. 1980. Monosaccharide composition of marine particles and sediments from the Bering Sea and northern North Pacific. *Ocean. Acta* **10**: 91–99.
- Thiele, S., B. M. Fuchs, R. Amann, and M. H. Iversen. 2015. Colonization in the photic zone and subsequent changes during sinking determine bacterial community composition in marine snow. *Appl. Environ. Microbiol.* **81**: 1463–1471. doi:10.1128/AEM.02570-14
- Wakeham, S. G., J. I. Hedges, C. Lee, M. L. Peterson, and P. J. Hernes. 1997. Compositions and transport of lipid biomarkers through the water column and surficial sediments of the equatorial Pacific Ocean. *Deep. Res. Part II Top. Stud. Oceanogr.* **44**: 2131–2162. doi:10.1016/S0967-0645(97)00035-0
- Wiedmann, I., M. Reigstad, A. Sundfjord, and S. Basedow. 2014. Potential drivers of sinking particle's size spectra and vertical flux of particulate organic carbon (POC): Turbulence, phytoplankton, and zooplankton. *J. Geophys. Res. Ocean.* **119**: 6900–6917. doi:10.1002/2013JC009305.Received
- Yancey, P. H., M. E. Clark, S. C. Hand, R. D. Bowlus, and G. N. Somero. 1982. Classes of Intracellular Osmolyte Systems and Their Distributions Living with Water Stress: Evolution of Osmolyte Systems. *Science* (80-.). **217**: 1214–1222. doi:10.1126/science.7112124

Table 1. Percent of samples at each water column depth and in the particle traps that contained each metabolite.

| Compound | Surface | DCM | 250 m | Particle traps |
|-----------------------------------------|----------------|------------|--------------|-----------------------|
| 1-deoxy-D-xylulose-5-phosphate | 13 | 9 | 13 | 17 |
| 2,3-dihydroxypropane-1-sulfonate (DHPS) | 25 | 9 | 0 | 58 |
| 4-hydroxybenzoic acid | 25 | 0 | 0 | 8 |
| 5-methylthioadenosine (MTA) | 50 | 55 | 0 | 100 |
| adenine | 38 | 18 | 38 | 8 |
| adenosine | 75 | 73 | 13 | 100 |
| adenosine 5'-monophosphate (AMP) | 75 | 73 | 38 | 33 |
| arginine | 25 | 18 | 38 | 33 |
| betaine | 75 | 91 | 38 | 100 |
| biotin | 25 | 0 | 0 | 17 |
| caffeine | 25 | 27 | 75 | 100 |
| chitobiose | 0 | 9 | 0 | 8 |
| citrulline | 38 | 0 | 0 | 42 |
| cyanocobalamin | 0 | 0 | 0 | 8 |
| cysteine | 13 | 0 | 0 | 0 |
| D-glucosamine 6-phosphate | 0 | 0 | 13 | 0 |
| D-ribose 5-phosphate | 13 | 9 | 13 | 0 |
| dimethylsulfoniopropionate (DMSP) | 100 | 100 | 38 | 100 |
| folic acid | 13 | 0 | 13 | 8 |
| fosfomycin | 0 | 0 | 0 | 8 |
| glucose 6-phosphate | 0 | 18 | 13 | 8 |
| glutamic acid | 50 | 55 | 0 | 58 |
| glutamine | 0 | 9 | 0 | 8 |
| guanine | 100 | 100 | 100 | 100 |
| indole 3-acetic acid (IAA) | 0 | 0 | 0 | 75 |
| inosine | 88 | 55 | 50 | 92 |
| isoleucine | 63 | 55 | 25 | 92 |
| methionine | 0 | 0 | 13 | 33 |
| <i>N</i> -acetylglutamic acid | 13 | 0 | 0 | 25 |
| ornithine | 25 | 18 | 0 | 0 |
| orotic acid | 0 | 18 | 0 | 0 |
| pantothenic acid | 0 | 0 | 0 | 8 |
| phenylalanine | 100 | 100 | 100 | 100 |
| proline | 75 | 45 | 25 | 92 |
| pyridoxine | 38 | 27 | 25 | 8 |
| riboflavin | 13 | 27 | 0 | 100 |
| <i>sn</i> -glycerol 3-phosphate | 0 | 9 | 0 | 33 |
| tryptophan | 75 | 91 | 63 | 92 |
| uracil | 25 | 27 | 0 | 58 |
| uridine 5'-monophosphate (UMP) | 0 | 9 | 13 | 25 |
| xanthine | 38 | 55 | 0 | 100 |
| 2,3-dihydroxybenzoic acid | 0 | 0 | 0 | 0 |

| | | | | |
|----------------------------------------------|---|---|---|---|
| 3-mercaptopropionate | 0 | 0 | 0 | 0 |
| 4-aminobenzoic acid | 0 | 0 | 0 | 0 |
| ciliatine | 0 | 0 | 0 | 0 |
| cytosine | 0 | 0 | 0 | 0 |
| desthiobiotin | 0 | 0 | 0 | 0 |
| dihydroxyacetone phosphate | 0 | 0 | 0 | 0 |
| glutathione | 0 | 0 | 0 | 0 |
| glyphosate | 0 | 0 | 0 | 0 |
| inosine 5'-monophosphate (IMP) | 0 | 0 | 0 | 0 |
| <i>N</i> -acetylglucosamine | 0 | 0 | 0 | 0 |
| <i>S</i> -(5'-adenosyl)-L-homocysteine (SAH) | 0 | 0 | 0 | 0 |
| sarcosine | 0 | 0 | 0 | 0 |
| taurine | 0 | 0 | 0 | 0 |
| threonine | 0 | 0 | 0 | 0 |
| thymidine | 0 | 0 | 0 | 0 |

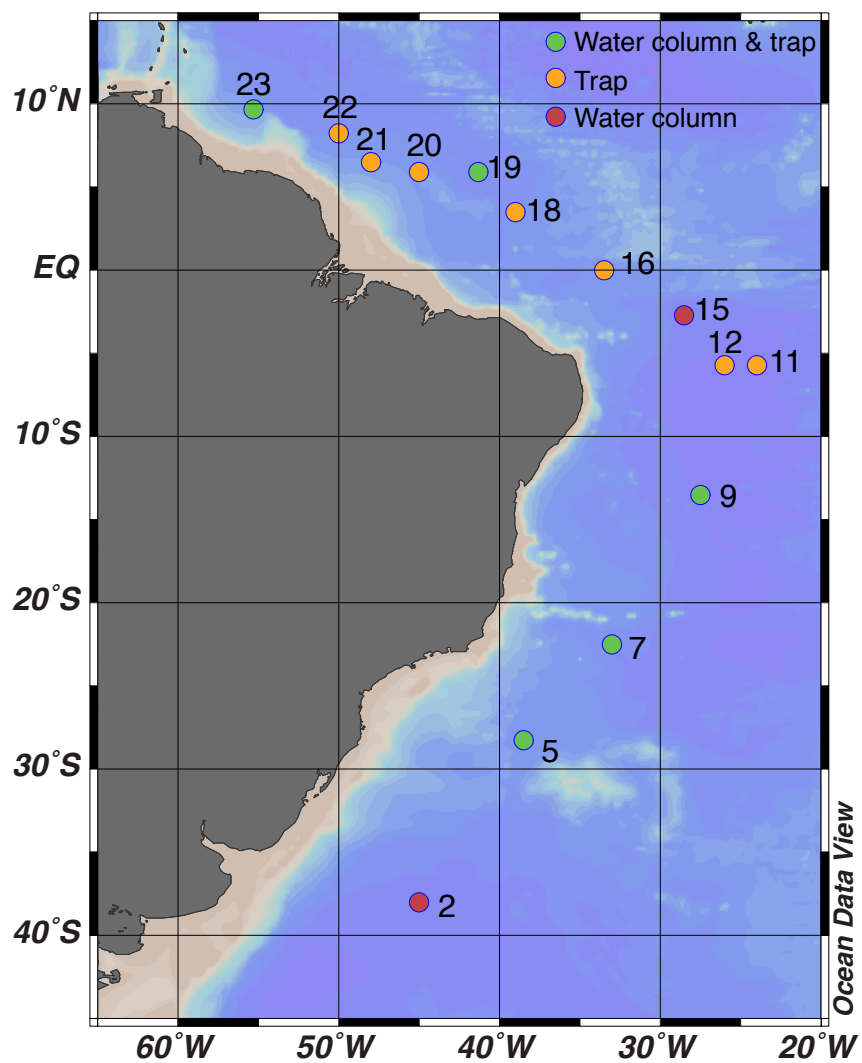


Figure 1. Map of transect. Stations are colored by what sample types are present for each.

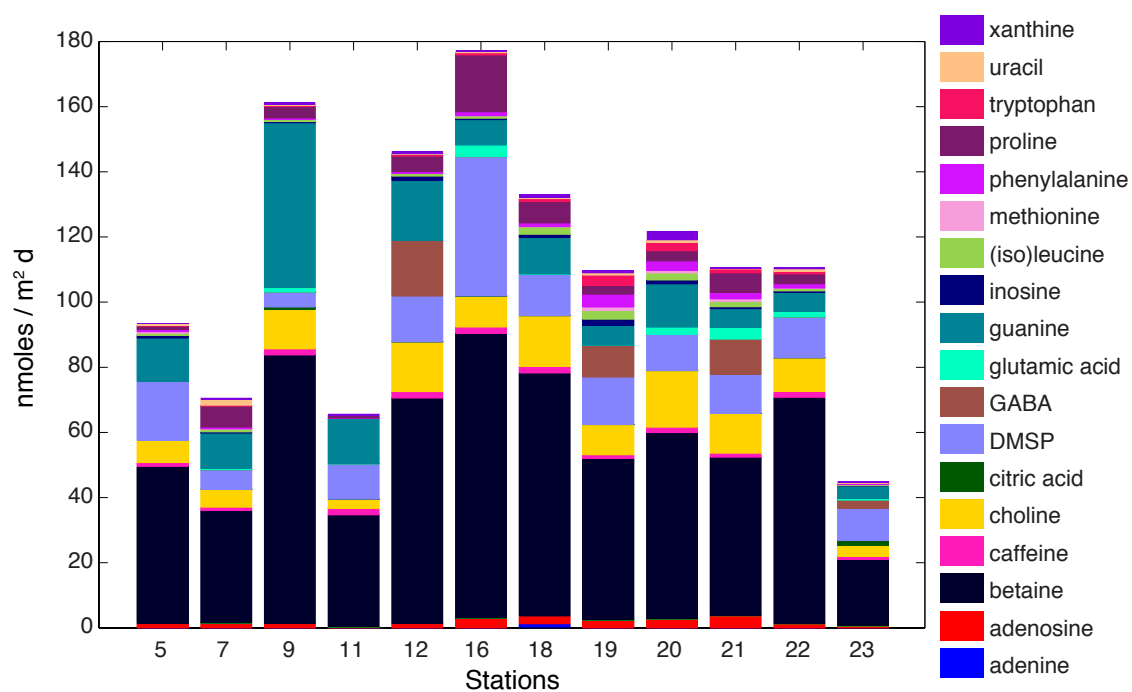


Figure 2. Fluxes of metabolites into the particle traps at 150 m. Only metabolites that had a flux greater 1 nmol m⁻² d⁻¹ in at least one sample are shown.

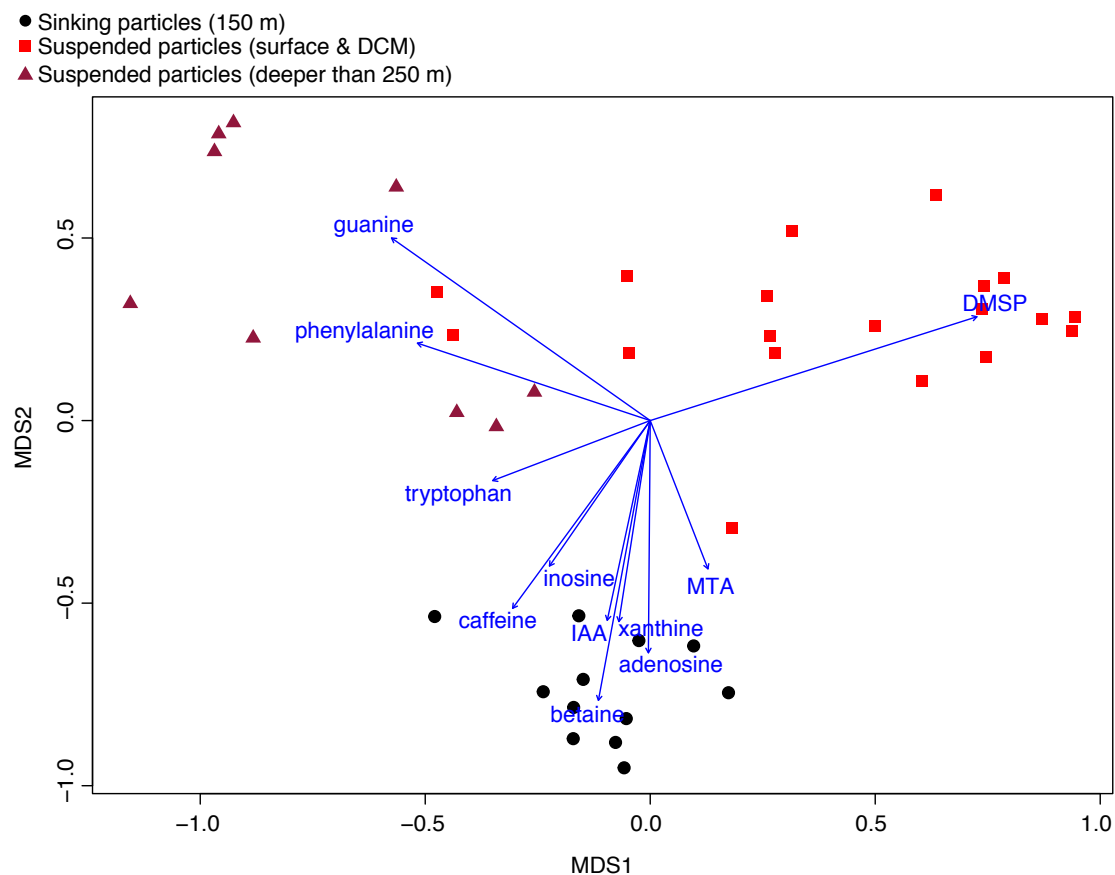


Figure 3. Non-metric multidimensional scaling shows that the metabolite compositions of the suspended particles are distinct from that of the sinking particles. Each marker represents a different sample and they are clustered by the relative composition of the metabolites in each sample. The length of the lines for each metabolite indicate how strong a relationship that metabolite has with the cluster of samples it points towards.

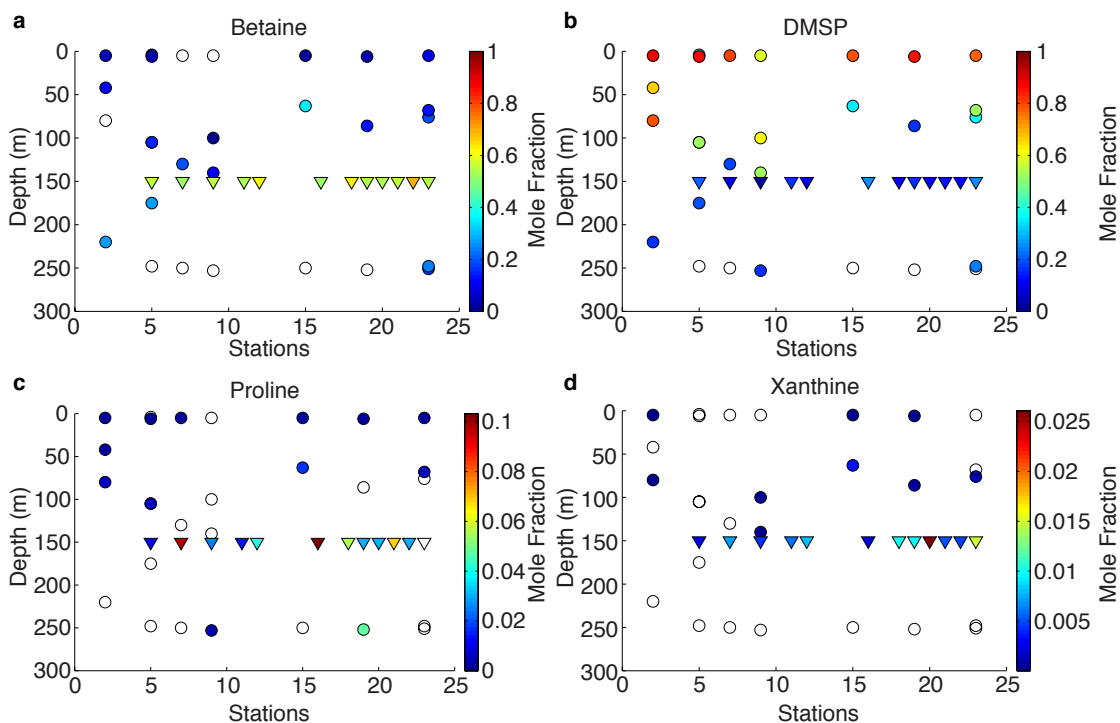


Figure 4. Metabolite profiles in both the suspended (circles) and sinking particles (triangles). Open symbols indicate samples where the metabolite was not detected. Warmer colors indicate a higher mole fraction of that metabolite (metabolites normalized to total moles targeted metabolites) in the sample. Shown are **a)** betaine, **b)** DMSP, **c)** proline, **d)** xanthine.

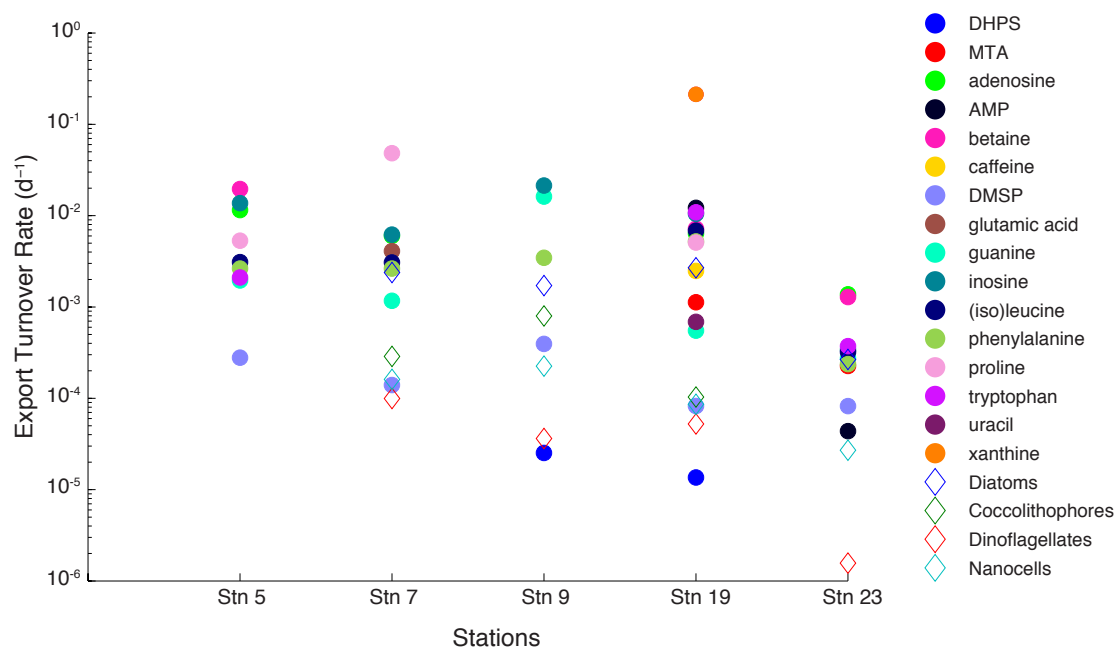


Figure 5. The export turnover rates for metabolites that were detected at both the surface and in the particle trap for a given station. The turnover rate reflects the fraction relative to the surface of metabolite that reaches 150 m. Also plotted are the export turnover rates of phytoplankton cells (obtained from Durkin et al. 2016).

Supplemental Information

Table S1. Deployment times of net traps, the mixed layer depth, and the POC flux for each station discussed. Data obtained from Durkin et al. (2016).

| Station | Deployment time (h) | Mixed layer depth (m) | POC flux ($\text{mg m}^{-2} \text{d}^{-1}$) |
|---------|---------------------|-----------------------|-----------------------------------------------|
| 5 | 25 | 70 | 51 ± 11 |
| 7 | 24 | 54 | 30 ± 12 |
| 9 | 15.5 | 48 | 38 ± 5 |
| 11 | 11.5 | 72 | 52 ± 6 |
| 12 | 13.5 | 68 | 37 ± 10 |
| 16 | 13 | 50 | 30 ± 10 |
| 18 | 14 | 60 | 45 ± 6 |
| 19 | 26 | 91 | 49 ± 4 |
| 20 | 17 | 100 | 52 ± 11 |
| 21 | 17.5 | 138 | 26 |
| 22 | 13.5 | 80 | 63 ± 14 |
| 23 | 26 | 82 | 43 |

Table S2. Metabolites that passed quality control checks in particle trap samples and the number of particle trap samples in which they were detected.

| Compound | Number of samples (n =12) |
|-----------------------------------------|----------------------------------|
| 1-deoxy-D-xylulose-5-phosphate | 2 |
| 2,3-dihydroxypropane-1-sulfonate (DHPS) | 7 |
| 4-hydroxybenzoic acid | 1 |
| 5-methylthioadenosine (MTA) | 12 |
| adenine | 1 |
| adenosine | 12 |
| adenosine 5'-monophosphate (AMP) | 4 |
| arginine | 4 |
| betaine | 12 |
| biotin | 2 |
| caffeine | 12 |
| chitobiose | 1 |
| choline | 12 |
| citric acid | 2 |
| citrulline | 5 |
| cyanocobalamin | 1 |
| dimethylsulfoniopropionate (DMSP) | 12 |
| folic acid | 1 |
| fosfomycin | 1 |
| gamma-aminobutyric acid (GABA) | 4 |
| glucose 6-phosphate | 1 |
| glutamic acid | 7 |
| glutamine | 1 |
| guanine | 12 |
| indole 3-acetic acid (IAA) | 9 |
| inosine | 11 |
| (iso)leucine | 11 |
| malic acid | 2 |
| methionine | 4 |
| <i>N</i> -acetylglutamic acid | 3 |
| pantothenic acid | 1 |
| phenylalanine | 12 |
| proline | 11 |
| pyridoxine | 1 |
| riboflavin | 12 |
| <i>sn</i> -glycerol 3-phosphate | 4 |
| succinic acid | 4 |
| taurocholate | 4 |
| tryptophan | 11 |
| uracil | 7 |
| uridine 5'-monophosphate (UMP) | 3 |
| xanthine | 12 |
| 2,3-dihydroxybenzoic acid | 0 |

| | |
|----------------------------------------------|---|
| 3-mercaptopropionate | 0 |
| 4-aminobenzoic acid | 0 |
| 6-phosphogluconic acid | 0 |
| acetyltaurine | 0 |
| chitotriose | 0 |
| ciliatine | 0 |
| cysteine | 0 |
| cytosine | 0 |
| D-glucosamine 6-phosphate | 0 |
| D-ribose 5-phosphate | 0 |
| desthiobiotin | 0 |
| dihydroxyacetone phosphate | 0 |
| glutathione | 0 |
| glyphosate | 0 |
| inosine 5'-monophosphate (IMP) | 0 |
| muramic acid | 0 |
| <i>N</i> -acetylglucosamine | 0 |
| NAD | 0 |
| NADP | 0 |
| ornithine | 0 |
| orotic acid | 0 |
| phosphoenolpyruvate | 0 |
| <i>S</i> -(5'-adenosyl)-L-homocysteine (SAH) | 0 |
| sarcosine | 0 |
| serine | 0 |
| taurine | 0 |
| thiamin monophosphate | 0 |
| threonine | 0 |
| thymidine | 0 |

Table S3. Values of export turnover rates (d^{-1}). These values are plotted in Figure 5. --- indicates when the export turnover rate of a metabolite could not be calculated.

| Compound | Station 5 | Station 7 | Station 9 | Station 19 | Station 23 |
|---------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| DHPS | --- | --- | 3×10^{-5} | 1×10^{-5} | --- |
| MTA | --- | --- | --- | 1×10^{-3} | 2×10^{-4} |
| adenosine | 1×10^{-2} | 6×10^{-3} | --- | 6×10^{-3} | 1×10^{-3} |
| AMP | 0 | --- | --- | 1×10^{-2} | 4×10^{-5} |
| betaine | 2×10^{-2} | --- | --- | 7×10^{-3} | 1×10^{-3} |
| caffeine | --- | --- | --- | 2×10^{-3} | --- |
| citrulline | --- | --- | --- | 8×10^{-4} | --- |
| DMSP | 3×10^{-4} | 1×10^{-4} | 4×10^{-4} | 8×10^{-5} | 8×10^{-5} |
| glutamic acid | --- | 4×10^{-3} | --- | 0 | --- |
| guanine | 2×10^{-3} | 1×10^{-3} | 2×10^{-2} | 6×10^{-4} | 3×10^{-4} |
| inosine | 1×10^{-2} | 6×10^{-3} | 2×10^{-2} | 1×10^{-2} | 3×10^{-4} |
| (iso)leucine | 3×10^{-3} | 3×10^{-3} | --- | 7×10^{-3} | 3×10^{-4} |
| phenylalanine | 3×10^{-3} | 3×10^{-3} | 3×10^{-3} | 5×10^{-3} | 2×10^{-4} |
| proline | 5×10^{-3} | 5×10^{-2} | --- | 5×10^{-3} | 0 |
| tryptophan | 2×10^{-3} | --- | --- | 1×10^{-2} | 4×10^{-4} |
| uracil | --- | --- | --- | 7×10^{-4} | --- |
| xanthine | --- | --- | --- | 2×10^{-1} | --- |

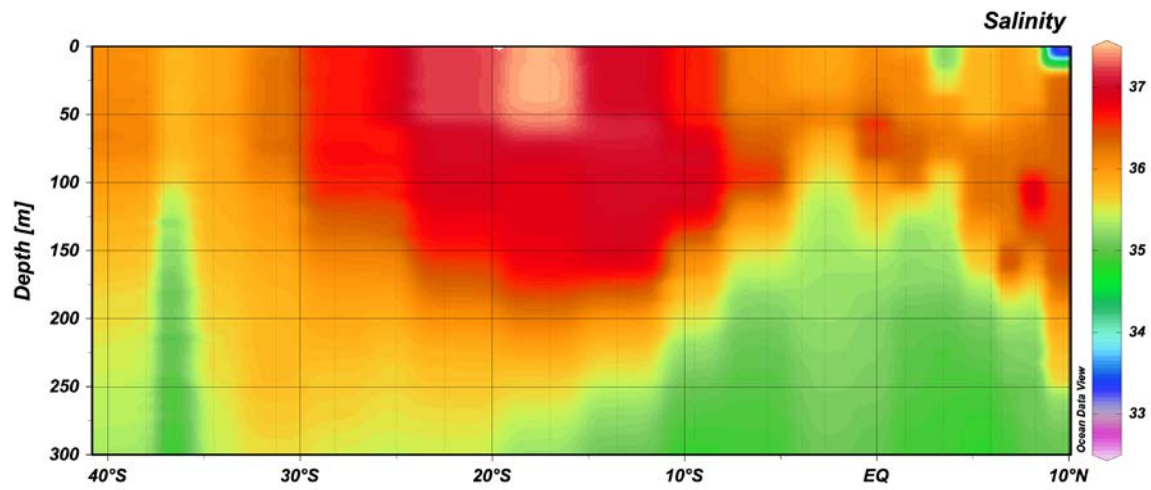


Figure S1. Salinity (ppt) profile along transect.

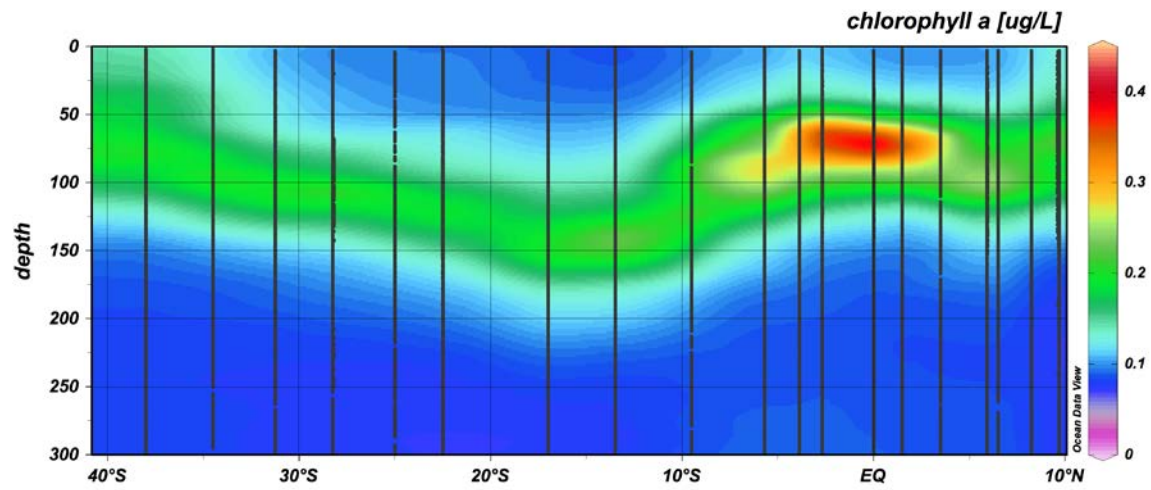


Figure S2. Chlorophyll *a* profile along the transect.

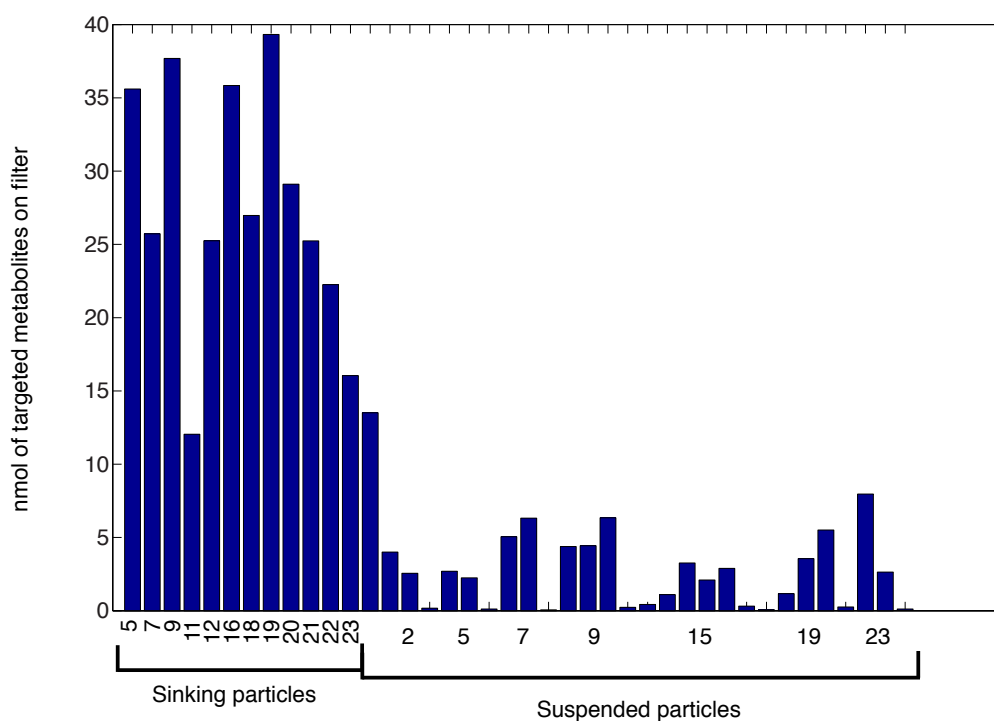


Figure S3. Total moles of particulate targeted metabolites collected on filters. A comparison of the sinking particles and the suspended particles indicates that larger amounts of targeted metabolites were measured on the sinking particles due to a larger amount of particles collected.

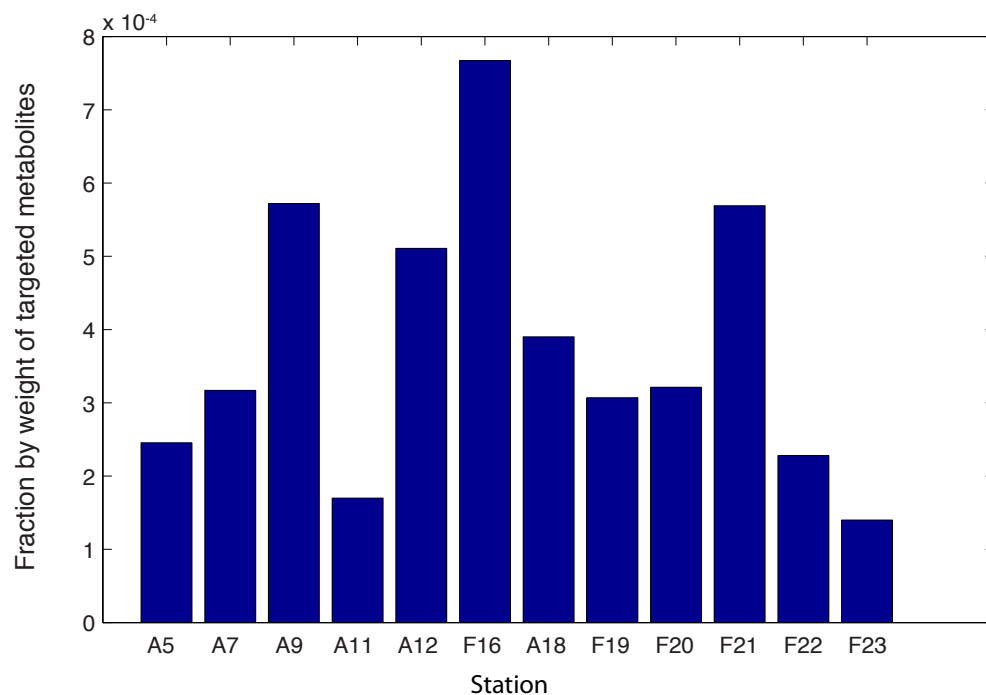


Figure S4. The fraction, by weight, of the flux of targeted metabolites measured relative to the total POC flux. *Note, the POC flux was measured during the same deployment but at 125 m while the metabolites were collected at 150 m.

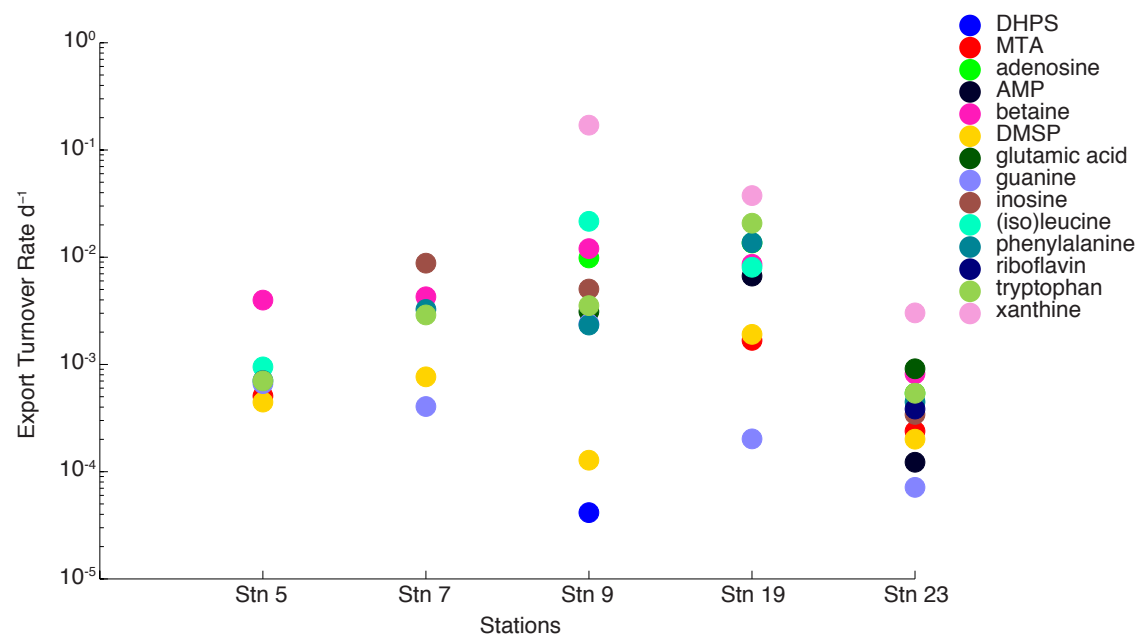


Figure S5. Export turnover rates (d^{-1}) calculated with the metabolite abundance at the DCM rather than the surface.

Chapter 6. Conclusions

6.1 Summary of major findings

This thesis advances our ability to study the role of metabolites in microbial ecology and the impact of microbial communities on organic matter cycling. The development and characterization of targeted metabolomics methods in Chapter 2 allows, in particular, more accurate measurements of dissolved metabolites extracted from seawater and a better understanding of the point at which higher DOC concentrations impact the ionization efficiency of metabolites. In Chapter 3 growth on an individual metabolite, DMSP, resulted in a cascade of metabolic changes in *R. pomeroyi*. This demonstrated that an individual metabolite in the ocean can have a significant impact on the metabolism of surrounding microbes. Thus, metabolite distributions and variability in the ocean need to be more extensively documented. The genetic capabilities of an organism must be studied under a range of conditions including the composition of the metabolites to which it is exposed. In Chapter 4, extensive characterization of a suite of metabolites in the western Atlantic Ocean provides the basis for understanding the variability of metabolites throughout oceanic regions. Identification of metabolites whose distributions can be linked to biomass versus those that seem to be responding to a more complex set of environmental factors can lead us to a deeper understanding of physiological adaptations in the marine environment and, in turn, the subsequent effects on metabolite availability in the microbial community. The fact that core metabolite composition, metabolites that are produced across the domains of life, is not the same everywhere in the ocean may have profound implications for community composition and succession. Chapter 5 explored metabolite dynamics on sinking particles and how the composition reflects both the source material in the euphotic zone as well as degradation processes during sinking. Overall, this thesis links

metabolite distributions within important reservoirs and fluxes of the marine carbon cycle including suspended particles, the dissolved pool, and sinking particles as well as examining a specific metabolite-microbe interaction that may be indicative of the shifts microbes undergo moving between a free-living lifestyle and a particle-associated one. All of this work lays the foundation for focused study on specific metabolites and a general understanding of metabolite variability on large spatial scales. This framework of targeted metabolite dynamics may also facilitate linking the discovery of novel or unexpected metabolites to already well-characterized components of metabolic pathways.

6.2 Metabolites for further study

From the work in this thesis, there are a variety of metabolites whose regulation and implications require further study to identify their role in metabolism in marine organisms and the controls on their production and degradation in the ocean. These metabolites may have implications for how marine microbes respond to their environment and give insight into the metabolisms responsible for organic matter cycling.

Osmolytes emerged as an important group of metabolites in the studies undertaken in this thesis. They are organic molecules that are present at high concentrations inside of cells resulting in a significant impact on the metabolite composition of marine waters and thus available carbon substrates for other members of the microbial community. In addition, the relative intracellular concentrations of osmolytes can respond to a variety of external factors. For example, DMSP is an osmolyte that can serve as an important carbon and sulfur source for marine bacteria. However, this work indicates that it may have farther-reaching consequences on microbial metabolism that must be further studied, as well as being degraded more rapidly than other metabolites on sinking particles. Other osmolytes, such as betaine, were found to be of perhaps

greater importance in the communities on sinking particles than in suspended water column.

Proline, also known to act as an osmolyte, was measured in the deep ocean as well as appearing to be exported from the euphotic zone on sinking particles. The extent of its importance as an osmolyte in deep ocean communities is not understood, however. The roles of these molecules are complex and merit further study.

Certain nucleic acid metabolites showed variability that could be indicative of environmental conditions or microbial processing of DNA and RNA. It is unclear why adenine shows no relationship to biomass while adenosine and AMP both do; this may be related to responses to nutrient limitation. Other derivatives like xanthine and uracil may be linked to degradation of nucleic acids or cell death but, again the regulation of their production in metabolism is incompletely understood and completely unstudied in the ocean. Guanine displays surprisingly high concentrations and distributions throughout the ocean and it is unclear why this purine nucleobase has extremely different distributions from adenine.

Other metabolic intermediates and vitamins demonstrate dynamics that could be instructive of how microbial communities function, as well as an individual microbe's differing metabolic strategies. For example, MTA was an indicator of AHL synthesis in *R. pomeroyi* cultures and also showed distinctive dissolved distributions in the ocean. Riboflavin and alpha-ribazole were both elevated in DMSP treatments in culture. When measured in the ocean, dissolved riboflavin was strongly linked to chlorophyll *a* concentrations. However, alpha-ribazole, a component of vitamin B₁₂, has not yet been measured in the ocean. These molecules could play important roles in microbial interactions, but we do not currently know the environmental and metabolic controls on their production.

6.3 Ecophysiology and marine metabolites

Part of the challenge of interpreting metabolite abundances in both culture experiments and in the field is our lack of a comprehensive understanding of the factors that control metabolite abundances inside of cells, particularly for marine microorganisms. For instance, glutamine displayed an intriguing response to *R. pomeroyi* growth on DMSP, but its role as a regulator of nitrogen acquisition has predominantly been studied in *E. coli*. In *E. coli*, changes in glutamine concentration have been observed in response to nitrogen limitation; in contrast, other factors that control glutamine abundance have not been extensively studied (Reitzer 2003; Chubukov et al. 2014). A full suite of environmental controls, from abiotic factors like temperature and light, to nutrient limitations, to food web interactions like grazing, viral infection, and mutualistic interactions, must all be examined for a variety of marine microbes. This will aid in the interpretation metabolite variability by linking it to specific environmental conditions. Ultimately, this metabolite variability contributes to differences in organic matter production as well as other microbial ecological functions.

6.4 How metabolites impact microbial ecology and carbon cycling

An intriguing possibility that arose from the *R. pomeroyi* experiment was the possible remodeling of metabolism in response to DMSP to increase utilization of organic substrates that would be available in an organic matter rich environment. This concept must be further tested, but if a single metabolite can change how a microbe interacts with the organic matter in its environment, individual metabolite distributions may have a greater impact on organic matter cycling than would be supposed based on their abundance alone. Additionally, the availability of certain metabolites in the ocean may shape microbial interactions in ways we have not yet documented. Other studies have found interactions like providing tryptophan for synthesis of indole-3-acetic acid (Amin et al. 2015), co-culturing resulting in increased production of amino

acids (Paul et al. 2012), or provision of a organosulfur substrate in exchange for vitamin B₁₂ (Durham et al. 2015). In addition, auxotrophy in marine microbes may be an important aspect of metabolite exchange, but it is challenging to study due to the difficulty of culturing these organisms. To observe and identify these interactions in the ocean it will be essential to quantify at what concentrations metabolites can be detected by microbes and transported into the cell. The role of metabolite distributions in shaping community structure due to specific metabolite requirements has not been extensively studied in the ocean at this point. As marine metabolomics methods improve and our map of ocean metabolite concentrations expands, an important aspect of understanding metabolite exchange and cycling will be to move from measurements of abundance to quantification of rates of production and removal of metabolites (Sauer 2006; Reaves and Rabinowitz 2011).

6.5 Linking metabolomics to other omics datasets

Moving forward, an essential way to accelerate our understanding of the controls on metabolite distributions and their impact on microbial communities and organic matter cycling is to link metabolite profiles to other omics datasets. Integrating these types of datasets presents a variety of challenges including the computational aspects of linking multiple large datasets with fundamentally different types of data. The tools to do this are being developed, but the number of areas of expertise that must be combined to make these types of comparisons possible requires large-scale collaborations and projects. As these tools advance, however, they hold great promise for allowing us to attribute certain metabolites to certain species and to link production and degradation of metabolites to specific pathways, some of which may be unknown at this point in time (Prosser et al. 2014). As we continue to expand our understanding of the relationship between metabolite distributions and species composition and activity, the ocean metabolome

may begin to be used in ways similar to those currently in applied in medicine. Metabolite biomarkers may be used to identify specific biochemical activities, ecosystem types, or even to diagnose changes in ocean health in response to a warming and acidifying ocean. To do this, there is a great deal left to discover, and this thesis provides a fundamental dataset from which to advance our understanding of the roles of metabolites in the ocean.

References

- Amin, S. A., L. R. Hmelo, H. M. van Tol, and others. 2015. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **522**: 98–101. doi:10.1038/nature14488
- Chubukov, V., L. Gerosa, K. Kochanowski, and U. Sauer. 2014. Coordination of microbial metabolism. *Nat. Rev. Microbiol.* **12**: 327–340. doi:10.1038/nrmicro3238
- Durham, B. P., S. Sharma, H. Luo, and others. 2015. Cryptic carbon and sulfur cycling between surface ocean plankton. *Proc. Natl. Acad. Sci. U. S. A.* **112**: 453–457. doi:10.1073/pnas.1413137112
- Paul, C., M. A. Mausz, and G. Pohnert. 2012. A co-culturing/metabolomics approach to investigate chemically mediated interactions of planktonic organisms reveals influence of bacteria on diatom metabolism. *Metabolomics* **9**: 349–359. doi:10.1007/s11306-012-0453-1
- Prosser, G. a, G. Larrouy-Maumus, and L. P. S. de Carvalho. 2014. Metabolomic strategies for the identification of new enzyme functions and metabolic pathways. *EMBO Rep.* **15**: 657–69. doi:10.15252/embr.201338283
- Reaves, M. L., and J. D. Rabinowitz. 2011. Metabolomics in systems microbiology. *Curr. Opin. Biotechnol.* **22**: 17–25. doi:10.1016/j.copbio.2010.10.001
- Reitzer, L. 2003. Nitrogen assimilation and global regulation in *Escherichia coli*. *Annu. Rev. Microbiol.* **57**: 155–176. doi:10.1146/annurev.micro.57.030502.090820
- Sauer, U. 2006. Metabolic networks in motion: ¹³C-based flux analysis. *Mol. Syst. Biol.* **2**: 62. doi:10.1038/msb4100109